

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 43.583 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802

Sequence: 1 MNSRRSLRWLSVLA.....QTKSGTRGNTVLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003ss:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1802	100.0	355	2	Aaw32367 Mycobacte
2	1802	100.0	355	2	Aaw32435 Mycobacte
3	1802	100.0	355	2	Aaw64307 Mycobacte
4	1802	100.0	355	2	Aaw81670 M. tuberc
5	1802	100.0	355	2	Aay38972 M. tuberc
6	1802	100.0	355	2	Aay39109 M. tuberc
7	1802	100.0	355	4	Aau01890 M. tuberc
8	1802	100.0	355	5	Aae29701 Mycobacte
9	1802	100.0	355	5	Aae17565 Mycobacte
10	1792	99.4	355	2	Aay05000 Mycobacte
11	1792	99.4	355	4	Aao22137 Mycobacte
12	1792	99.4	355	4	Aag81110 Mycobacte
13	1792	99.4	379	2	Aay04830 Mycobacte
14	1634	90.7	330	5	Aae29702 Mycobacte
15	1634	90.7	330	5	Aae17566 Mycobacte
16	1631	90.5	330	5	Aae29703 Mycobacte
17	1631	90.5	330	5	Aae17567 Mycobacte
18	1631	90.5	330	7	ADA26372 Mycobacte
19	1631	90.5	723	7	ADA26354 Mycobacte
20	1631	90.5	1010	7	ADA26356 Mycobacte
21	1627	90.3	330	7	ADA26371 Mycobacte
22	999.5	55.5	1016	7	ADA26370 M. bovis
23	992	55.0	1154	7	ADA26368 Mycobacte
24	990.5	55.0	825	7	ADA26366 Mycobacte
25	989.5	54.9	813	7	ADA26367 Mycobacte

26 988 54.8 1022 7 ADA26369 Mycobacte
27 985 54.7 195 5 AAE29704 Mycobacte
28 985 54.7 195 5 AAE17568 Mycobacte
29 983.5 54.6 875 7 ADA26365 Mycobacte
30 983 54.6 596 2 AAY32070 Mycobacte
31 983 54.6 596 5 AAE29710 Mycobacte
32 983 54.6 596 5 AAE17574 Mycobacte
33 983 54.6 599 5 AAU74599 Antigenic
34 983 54.6 729 4 AAO22142 Ral12-H9-3
35 983 54.6 729 5 AAE29708 Mycobacte
36 983 54.6 729 5 AAE17572 Mycobacte
37 983 54.6 729 7 ADA26373 Mycobacte
38 983 54.6 930 5 AAE29731 Mycobacte
39 983 54.6 930 7 ADA26364 Mycobacte
40 980 54.4 729 5 AAE29709 Mycobacte
41 980 54.4 729 5 AAE17573 Mycobacte
42 980 54.4 729 7 ADA26374 Mycobacte
43 978 54.3 726 5 AAU74588 Antigenic
44 978 54.3 729 2 AAY32059 Mycobacte
45 676 37.5 132 2 AAW32354 Mycobacte

ALIGNMENTS

RESULT 1
AAW32367
ID AAW32367 standard; protein; 355 AA.
AC AAW32367;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TBra35.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WC9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-MAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX
(CORI-) CORIXA CORP.
Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
Vedvick TH, Twardzik DR;
WPI; 1997-192904/17.
N-PSDB; AAT91414.
New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
useful for diagnosis of M. tuberculosis infection.
Example 3; Page 124-126; 190pp; English.
A new immunogenic polypeptide has been developed comprising an
immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
variant differing only in conservative substitutions and/or
modifications). The present sequence represents a M.tuberculosis antigen,
TBra35. The immunogenic polypeptide can be used to diagnose
M.tuberculosis infection by forming complexes with specific antibodies in
the sample. Fragments of DNA encoding the immunogenic polypeptide can be
used as diagnostic primers or probes and agents that bind to the antigen,

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis
 XX
 SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSSLLSLVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MNSRRSLRWSSLLSLVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60

QY 61 PQVNVINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFVSGQTYGVVDVVG 120
 DB 61 PQVNVINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFVSGQTYGVVDVVG 120

QY 121 YDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180

QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSGQGGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSGQGGA 240

QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAPAASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAPAASLGIST 300

QY 301 GDVITAVDGAIPINSATAMADALNGHPGDVISVNMWTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAIPINSATAMADALNGHPGDVISVNMWTKSGGTRTGNVTLAEGPPA 355

RESULT 2
 AAW32435
 ID AAW32435 standard; protein; 355 AA.
 AC AAW32435;
 XX
 DT 08-JAN-1998 (first entry)
 DE Mycobacterium tuberculosis antigen TbrA35.
 XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.
 XX Mycobacterium tuberculosis.
 OS WO9709428-A2.
 PN 13-MAR-1997.
 XX
 PF 30-AUG-1996; 96WO-US014674.
 XX
 PR 01-SEP-1995; 95US-00523436.
 PR 22-SEP-1995; 95US-00533634.
 PR 22-MAR-1996; 96US-00620874.
 PR 05-JUN-1996; 96US-00659683.
 PR 12-JUL-1996; 96US-00680574.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Read SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TH, Twardzik DR;
 XX
 DR WPI; 1997-192903/17.
 DR N-PSDB; AAT91477.
 XX
 XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.
 XX

PS Example 3; Page 114-116; 168pp; English.
 XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbrA35. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)
 XX
 SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSSLLSLVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MNSRRSLRWSSLLSLVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60

QY 61 PQVNVINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFVSGQTYGVVDVVG 120
 DB 61 PQVNVINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFVSGQTYGVVDVVG 120

QY 121 YDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180

QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSGQGGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSGQGGA 240

QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAPAASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAPAASLGIST 300

QY 301 GDVITAVDGAIPINSATAMADALNGHPGDVISVNMWTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAIPINSATAMADALNGHPGDVISVNMWTKSGGTRTGNVTLAEGPPA 355

RESULT 3
 AAW64307
 ID AAW64307 standard; protein; 355 AA.
 AC AAW64307;
 XX
 DT 17-OCT-2003 (revised)
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TbrA35.
 XX Tuberculosis; infection; diagnosis; antigen; TbrA35.
 XX Mycobacterium tuberculosis; strain H37Ra.
 OS WO9816645-A2.
 PN 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018214.
 XX
 PR 11-OCT-1996; 96US-00729622.
 PR 13-MAR-1997; 97US-00818111.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 DR WPI; 1998-251292/22.
 XX

```

DR N-PSDB; AAV44355.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
PT develop products for the detection of M. tuberculosis infection and
PT diagnosis of tuberculosis.
XX
XX Example 3; Page 115-116; 250pp; English.
XX
CC This polypeptide comprises Mycobacterium tuberculosis soluble antigen
CC TBRa35. It is encoded by a DNA sequence (see AAV44355) isolated from a M.
CC tuberculosis strain H37Ra expression library with rabbit anti-sera raised
CC against M. tuberculosis supernatant. No significant homology was found
CC between TBRa35 and Genebank database sequences. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAV44291-464379) comprising an antigenic portion of a
CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic kits for
CC detecting M. tuberculosis infection in a patient using the above
CC polypeptides, antibodies or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
CC field)
XX
SQ Sequence 355 AA;
Query Match 100.0%; Score 1802; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNVNTKLYNNVAVGAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
DB 61 PQVNVNTKLYNNVAVGAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY 121 YDRTQDVAVLQIRGAGLPSAAIGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQIRGAGLPSAAIGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFOLSQGGQGFA 240
DB 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFOLSQGGQGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPASLGIST 300
DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPASLGIST 300
QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAIPINSATAMADALNGHHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
RESULT 4
AAW81670
ID AAW81670 standard; protein; 355 AA.
XX
AC AAW81670;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide TBRa35.
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9816646-A2.
XX
PD 23-APR-1998.

```

```

XX 07-OCT-1997; 97WO-US018293.
XX
XX 11-OCT-1996; 96US-00730510.
XX
XX 13-MAR-1997; 97US-00818112.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Meto A, Houghton R;
XX Vedwick TS, Twardzik DR, Lodes MJ;
XX
XX WPI; 1998-261042/23.
XX
XX N-PSDB; AAV64463.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
XX develop products for the detection of M. tuberculosis infection and for
XX diagnosis, treatment and prevention of tuberculosis.
XX
XX Example 3; Page 110-111; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
XX Mycobacterium tuberculosis (MT) antigen which can be used in a method for
XX inducing protective immunity against tuberculosis (TB). This sequence can
XX be formulated into vaccines and/or pharmaceutical compositions for
XX immunising against M. tuberculosis infection or may be used for the
XX diagnosis of tuberculosis
XX
XX Sequence 355 AA;
Query Match 100.0%; Score 1802; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNVNTKLYNNVAVGAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
DB 61 PQVNVNTKLYNNVAVGAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY 121 YDRTQDVAVLQIRGAGLPSAAIGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQIRGAGLPSAAIGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFOLSQGGQGFA 240
DB 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFOLSQGGQGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPASLGIST 300
DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPASLGIST 300
QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAIPINSATAMADALNGHHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
RESULT 5
AAV38972
ID AAV38972 standard; protein; 355 AA.
XX
AC AAV38972;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis recombinant antigen protein TBRa35.
XX
XX Antigen; diagnosis; detection; infection; antibody; immunisation;
XX vaccine; immunity.
XX
OS Mycobacterium tuberculosis.
XX

```

PN	WO9942118-A2.	XX	OS	Mycobacterium tuberculosis.	XX
PD	26-AUG-1999.	XX	PN	WO9942076-A2.	XX
PF	17-FEB-1999; 99WO-US003265.	XX	PD	26-AUG-1999.	XX
PR	18-FEB-1998; 98US-00024753.	XX	PF	17-FEB-1999; 99WO-US003268.	XX
PR	05-MAY-1998; 98US-00072596.	XX	PR	18-FEB-1998; 98US-00025197.	XX
XX	(CORI-) CORIXA CORP.	XX	PR	05-MAY-1998; 98US-00072967.	XX
PA	Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;	XX	XX	(CORI-) CORIXA CORP.	XX
PI	Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;	XX	XX	Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;	XX
XX	WPI; 1999-527416/44.	XX	PI	Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;	XX
DR	N-PSDB; AAZ19053.	XX	PI	WPI; 1999-527409/44.	XX
XX	New polypeptide comprising antigenic portions of M. tuberculosis.	XX	XX	N-PSDB; AAZ19265.	XX
XX	Example 3; Page 151-153; 323pp; English.	XX	XX	New antigens from Mycobacterium tuberculosis useful in diagnostic skin	XX
XX	This invention describes novel recombinant antigens and their encoding	XX	XX	tests and protective or therapeutic vaccines or compositions.	XX
CC	nucleic acids derived from Mycobacterium tuberculosis. The novel	XX	XX	Example 3; Page 108-109; 299pp; English.	XX
CC	polypeptides are useful for detecting M. tuberculosis infection in a	XX	XX	The present invention describes polypeptides comprising an immunogenic	XX
CC	biological sample by detecting antibodies which bind with the	XX	CC	part of a Mycobacterium tuberculosis antigen (Ag). Also described are	XX
CC	polypeptides, and are useful as vaccines for immunizing against M.	XX	CC	vaccines and fusion protein containing M. tuberculosis Ag's. M.	XX
CC	tuberculosis infection. The new detection methods are needed as current	XX	CC	tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other	XX
CC	vaccination strategies do not provide 100% immunity	XX	CC	polypeptides fragments, can be used in pharmaceutical compositions or	XX
XX	Sequence 355 AA;	XX	CC	vaccines to generate a protective or therapeutic immune response to M.	XX
XX	Query Match 100.0%; Score 1802; DB 2; Length 355;	XX	CC	tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.	XX
XX	Best Local Similarity 100.0%; Pred. No. 7.2e-126;	XX	CC	Ag can induce proliferation of, or cytokine secretion by, T, B or natural	XX
XX	Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX	CC	killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249	XX
QY	1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60	QY	CC	to AAZ19460 and AAZ39083 to AAZ39225 are used in the exemplification of	XX
Db	1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60	Db	CC	the present invention	XX
QY	61 PQVNVNITKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120	QY	XX	Sequence 355 AA;	XX
Db	61 PQVNVNITKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120	XX	Query Match 100.0%; Score 1802; DB 2; Length 355;		
QY	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGQGTTPRAVPGRVVVGQTV 180	XX	Best Local Similarity 100.0%; Pred. No. 7.2e-126;		
Db	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGQGTTPRAVPGRVVVGQTV 180	XX	Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLVGMNTAASDNFQLSGGGGGA 240	QY	1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60		
Db	181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLVGMNTAASDNFQLSGGGGGA 240	Db	1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60		
QY	241 IPIQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300	QY	61 PQVNVNITKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120		
Db	241 IPIQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300	Db	61 PQVNVNITKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120		
QY	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355	QY	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGQGTTPRAVPGRVVVGQTV 180		
Db	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355	Db	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGQGTTPRAVPGRVVVGQTV 180		
RESULT 6		QY	181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLVGMNTAASDNFQLSGGGGGA 240		
AAZ39109		Db	181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLVGMNTAASDNFQLSGGGGGA 240		
ID	AAZ39109 standard; protein; 355 AA.	QY	241 IPIQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300		
XX		Db	241 IPIQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300		
AC	AAZ39109;	QY	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355		
XX		Db	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355		
DT	05-NOV-1999 (first entry)	RESULT 7			
XX		AAU01890			
DE	M. tuberculosis antigen TBra35 amino acid sequence.	ID	AAU01890 standard; protein; 355 AA.		
XX		XX			
KW	Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;	XX			
KW	immunotherapy; diagnosis; immunisation; vaccine; infection;	AC	AAU01890;		
KW	immune response; skin test.				

XX 29-AUG-2001 (first entry)
 XX M. tuberculosis antigen TbrA35 (Mtb32A).
 DE TbrA35; Mtb32A; antigen; vaccine; tuberculosis; AIDS;
 XX acquired immunodeficiency disease.
 KW Mycobacterium tuberculosis.
 XX WO200124820-A1.
 OS 12-APR-2001.
 XX 10-OCT-2000; 2000WO-US028095.
 XX 07-OCT-1999; 99US-0158338P.
 PR 07-OCT-1999; 99US-0158425P.
 XX (CORI-) CORIXA CORP.
 PA Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 PI WPI; 2001-290576/30.
 XX N-PSDB; AAS03781.
 DR Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens.
 XX Disclosure; Page 153-154; 168pp; English.
 XX The sequence represents Mycobacterium tuberculosis TbrA35 (also known as
 CC Mtb32A), an M. tuberculosis antigen. Compositions comprising at least 2
 CC heterologous antigens, as a fusion protein, and vectors expressing the
 CC fusion proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS
 XX Sequence 355 AA;
 SQ
 Query Match 100.0%; Score 1802; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNTNKLGYNNVAGAGTGIVIDPVGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 DB 61 PQVNTNKLGYNNVAGAGTGIVIDPVGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 QY 121 YDRTQDVAVLQLRGAGLPSAATGGGVAVGEPVAVMGNSGGGGTTPRAVPGRWALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGLPSAATGGGVAVGEPVAVMGNSGGGGTTPRAVPGRWALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGPA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGPA 240
 QY 241 IPFGAMAIAGQIRSGGSGPTVHIGFTAFGLGVNDNNGARVQRVWGSAPAAISLIGST 300
 DB 241 IPFGAMAIAGQIRSGGSGPTVHIGFTAFGLGVNDNNGARVQRVWGSAPAAISLIGST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 8

AAE29701
 ID AAE29701 standard; protein; 355 AA.
 XX AAE29701;
 AC AAE29701;
 XX 27-JAN-2003 (first entry)
 DT Mycobacterium tuberculosis Mtb32A antigenic protein.
 DE Vaccine; immunity; diagnostic agent; gene therapy; Mtb32A antigen;
 XX Ra35FL.
 KW Mycobacterium tuberculosis.
 XX WO200272792-A2.
 XX 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US008223.
 PF 13-MAR-2001; 2001US-0275837P.
 PR (CORI-) CORIXA CORP.
 PA Skeiky Y, Brannon M, Guderian J;
 PI WPI; 2002-759844/82.
 XX N-PSDB; AAD47076.
 PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 XX tuberculosis.
 PS Disclosure; Page 78-79; 155pp; English.
 XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis Mtb32A antigenic protein. Mtb32A is also referred to as
 XX Ra35FL
 SQ Sequence 355 AA;
 Query Match 100.0%; Score 1802; DB 5; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNTNKLGYNNVAGAGTGIVIDPVGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 DB 61 PQVNTNKLGYNNVAGAGTGIVIDPVGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 QY 121 YDRTQDVAVLQLRGAGLPSAATGGGVAVGEPVAVMGNSGGGGTTPRAVPGRWALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGLPSAATGGGVAVGEPVAVMGNSGGGGTTPRAVPGRWALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGPA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGPA 240

QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGARVQVVGSAAPAASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGARVQVVGSAAPAASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 9
 AAEL17565
 ID AAEL17565 standard; protein; 355 AA.
 AC AAEL17565;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium species MTB32A (Ra35FL) protein.
 XX
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
 XX
 OS Mycobacterium sp.
 XX
 PN W0200198460-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001WO-US019959.
 XX
 PR 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Alderson M;
 XX
 DR WPI; 2002-147798/19.
 DR N-PSDB; AAD28335.
 XX
 PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 PS Claim 70; Page 94-95; 136pp; English.
 XX
 CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) protein
 XX
 SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 5; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNRRRLRWSLLSVLAAYGLGLATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNRRRLRWSLLSVLAAYGLGLATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNVNTKLYNNNAVAGAGTGIVTDPNGVLTNNHVIAGATDINAFSVGSGQTVGVDDVVG 120
 DB 61 PQVNVNTKLYNNNAVAGAGTGIVTDPNGVLTNNHVIAGATDINAFSVGSGQTVGVDDVVG 120
 QY 121 YDRTQDVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDLSLTGAETLNLQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSGGGGGFA 240
 DB 181 QASDLSLTGAETLNLQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSGGGGGFA 240
 QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGARVQVVGSAAPAASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGARVQVVGSAAPAASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 10
 AAY05000
 ID AAY05000 standard; protein; 355 AA.
 AC AAY05000;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species protein sequence 50D.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 OS Mycobacterium sp.
 XX
 PN W09909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR001813.
 XX
 PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 XX
 DR WPI; 1999-181045/15.
 DR N-PSDB; AAX34251.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 PS Claim 32; Fig 50D; 309pp; French.
 XX
 CC Sequences AAY0742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 SQ Sequence 355 AA;

Query Match 99.4%; Score 1792; DB 2; Length 355;

	Best Local Similarity	99.4%; Pred. No. 4e-125;	Mismatches	2; Indels	0; Gaps	0;
	Matches	353; Conservative	0;			
Qy	1	MSNSRRSLRWSLLSVLAAVGLGLATAPQAAPPALSQDRFADFPAFLPLDPSAMVAQVA	60			
Dd	1	MSNSRRSLRWSLLSVLAAVGLGLATAPQAAPPALSQDRFADFPAFLPLDPSAMVAQG	60			
Qy	61	PQVVNIINTKLGYNNAVAGGTGIVIDPNGVLTNNHVIAGATTINAFVSQSQTQVDVVG	120			
Dd	61	PQVVNIINTKLGYNNAVAGGTGIVIDPNGVLTNNHVIAGATTINAFVSQSQTQVDVVG	120			
Qy	121	YDRTDQDVAVQLQRAGGLPSAAIGGGVAVGPEPVVAMGNSGGCGTTPRAVPGRVVALGQTV	180			
Dd	121	YDRTDQDVAVQLQRAGGLPSAAIGGGVAVGPEPVVAMGNSGGCGTTPRAVPGRVVALGQTV	180			
Qy	181	QASDSLFGABETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAASNLFQSGGGGFA	240			
Dd	181	QASDSLFGABETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAASNLFQSGGGGFA	240			
Qy	241	IPTGQMAIAIAGILRSGGSPVPIHGPTAFLGLGVVDNNGCARVQRVVVGSAPAASLGIST	300			
Dd	241	IPTGQMAIAIAGILRSGGSPVPIHGPTAFLGLGVVDNNGCARVQRVVVGSAPAASLGIST	300			
Qy	301	GDVITAVDGAPINSATAMADALNHHPGDVISVNWQTKSGGTRTGNTVLAEPPA	355			
Dd	301	GDVTAYVDGAPINSATAMADALNHHPGDVITSWTOKSGGTRTGNTVLAEPPA	355			

RESULT 11	
AAO22137	
ID	AAO22137 standard; protein; 355 AA.
XX	
XX	
AC	AAO22137;
XX	
XX	
DT	03-OCT-2002 (first entry)
XX	
DE	Mycobacterium tuberculosis MTB32A protein.
XX	
KW	Ra12; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
KW	vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
KW	immunogen; cytokine.

CC	antibodies to <i>M. tuberculosis</i> antigens in an individual indicates that
CC	the individual is infected with it. The fusion polypeptides are useful as
CC	sources of proteins for monitoring binding of serum antibodies to fusion
CC	proteins and as an immunogen to induce and/or enhance immune responses.
CC	The coding sequences can be ligated with a coding sequence of another
CC	molecule such as <i>M. tuberculosis</i> antigens, a cytokine or an adjuvant and
CC	can be used in vivo as a DNA vaccine. This sequence represents the
CC	<i>Mycobacterium tuberculosis</i> MTE32A protein
XX	
XX	Sequence 355 AA;
XX	
XX	Query Match 99.4%; Score 1792; DB 4; Length 355;
XX	Best Local Similarity 99.4%; Pred. No. 4e-125;
XX	Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 MSNSRRSLRWSLLSVLLAAVGLGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
DB	1 MSNSRRSLRWSLLSVLLAAVGLGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVG 60
QY	61 PQVYNINTKLYGNNAVGAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
DB	61 PQVYNINTKLYGNNAVGAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY	121 YDRQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVMGNSGGQGTFRVAVPGRVVALGQTV 180
DB	121 YDRQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVMGNSGGQGTFRVAVPGRVVALGQTV 180
QY	181 QASDSLTCGAETLNGLIQFDAAIQPGSDGGPVVNGLGQVWGMNTAASDNFQLSGQGQGPPA 240
DB	181 QASDSLTCGAETLNGLIQFDAAIQPGSDGGPVVNGLGQVWGMNTAASDNFQLSGQGQGPPA 240
QY	241 IPIQCAIMAIAGIIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVORVVVGSAPAAASLGIST 300
DB	241 IPIQCAIMAIAGIIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVORVVVGSAPAAASLGIST 300
QY	301 GDVITAVDVGAPINSATAMADALNGHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
DB	301 GDVITAVDVGAPINSATAMADALNGHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
XX	
XX	RESULT 12
XX	AAG81110
XX	ID AAG81110 standard; protein; 355 AA.
XX	AC AAG81110;
XX	DT 04-SEP-2001 (first entry)
XX	DE <i>Mycobacterium tuberculosis</i> potential drug target protein SEQ ID 161.
XX	Drug target; growth; organism viability; characterisation.
XX	<i>Mycobacterium tuberculosis</i> .

FN	WO200135317-A1.	
XX		
XX	17-MAY-2001.	
XX		
XX	13-NOV-2000; 200CWO-US0311152.	
XX		
XX	12-NOV-1999; 99US-0165086P.	
PR	12-NOV-1999; 99US-0165134P.	
PR	01-FEB-2000; 2000US-0179531P.	
XX		
XX	(REGC) UNIV CALIFORNIA.	
PA		
XX	Eisenberg D, Rotstein SH, Marcotte EM;	
PI		
XX		
XX	WPI; 2001-329193/34.	
DR	N-PSDB; AAHS1961.	
DR		
XX		
pt	identifying nucleotide or polypeptide sequence for use as drug target,	
pt	involves providing algorithm that analyzes a functional relationship	

PT between nucleotide or polypeptide sequences, and comparing the sequences.
 XX Disclosure; Page 157; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or
 CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism

XX Sequence 355 AA;

Query Match 99.4%; Score 1792; DB 4; Length 355;
 Best Local Similarity 99.4%; Pred. No. 4e-125;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPACAAPPALSQDRFADFPALPLDPSAMVAQVA 60
 Db 1 MSNSRRSLRWSLLSVLAAGVGLATAPACAAPPALSQDRFADFPALPLDPSAMVAQVG 60
 QY 61 PQVYNINIKLGYNNAVGAGTGIVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 Db 61 PQVYNINIKLGYNNAVGAGTGIVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 QY 121 YDRQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 Db 121 YDRQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGQGFA 240
 Db 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGQGFA 240
 QY 241 IPIQAMAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGNGARVQRVVGSAPAAASLGIST 300
 Db 241 IPIQAMAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGNGARVQRVVGSAPAAASLGIST 300
 QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 13
 AAY04830
 ID AAY04830 standard; protein; 379 AA.

XX AAY04830;

XX 06-JUL-1999 (first entry)

XX Mycobacterium species protein sequence 50F.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.

XX Mycobacterium sp.

XX WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR001913.

XX 14-AUG-1997; 97FR-00010404.

XX 11-SEP-1997; 97FR-00011325.

PA (INSP) INST PASTEUR.

XX Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;

PI Goguet De La Salmoniere Y;

XX WPI; 1999-181045/15.

DR N-PSDB; AAX34252.

XX Mycobacterial DNA vectors containing reporter constructs - for
 CC identifying coding or promoter sequences involved in infection-associated
 CC protein expression.

XX Claim 32; Fig 50F; 309pp; French.

XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection

XX Sequence 379 AA;

Query Match 99.4%; Score 1792; DB 2; Length 379;
 Best Local Similarity 99.4%; Pred. No. 4.3e-125;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPACAAPPALSQDRFADFPALPLDPSAMVAQVA 60
 Db 25 MSNSRRSLRWSLLSVLAAGVGLATAPACAAPPALSQDRFADFPALPLDPSAMVAQVG 84
 QY 61 PQVYNINIKLGYNNAVGAGTGIVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 Db 85 PQVYNINIKLGYNNAVGAGTGIVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 144
 QY 121 YDRQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 Db 145 YDRQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 204
 QY 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGQGFA 240
 Db 205 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGQGFA 264
 QY 241 IPIQAMAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGNGARVQRVVGSAPAAASLGIST 300
 Db 265 IPIQAMAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGNGARVQRVVGSAPAAASLGIST 324
 QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 Db 325 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 379

RESULT 14

AAE29702

ID AAE29702 standard; protein; 330 AA.

XX AAE29702;

XX 27-JAN-2003 (first entry)

XX Mycobacterium tuberculosis mature Ra35 antigenic protein.

XX Vaccine; immunity; diagnostic agent; gene therapy; Ra35 antigen.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Misc-difference 182 /note= "Encoded by GAG"

FT Misc-difference 183 /note= "Encoded by GCG"

XX WO200272792-A2.

XX 19-SEP-2002.
XX
XX 13-MAR-2002; 2002WO-US008223.
XX
XX 13-MAR-2001; 2001US-0275637P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Brannon M, Guderian J;
XX
XX WPI; 2002-759844/82.
XX N-PSDB; AAD47077.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX tuberculosis.
XX
XX Disclosure; Page 79-80; 155pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a
XX fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX polynucleotide sequence encoding an antigen or an antigenic fragment from
XX Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX polypeptide or its fragment. The Leishmania polynucleotide is selected
XX from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
XX are used in methods for eliciting immune response in mammals. They are
XX useful as vaccines to elicit protective immunity against pathogenic
XX microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX polypeptides are used for enhancing the expression of polynucleotides, as
XX in vivo diagnostic agents and for raising antibodies in a non-human
XX animal. The invention is used in gene therapy. The present sequence is M.
XX tuberculosis mature Ra35 (N-terminus of MTB32A; RA35FL) antigenic protein
XX
XX Sequence 330 AA;
XX
Query Match 90.7%; Score 1634; DB 5; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.1e-113;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNAVGAGTGIVDPNGVILT 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNAVGAGTGIVDPNGVILT 67
QY 93 NNHVIAGATDINAFSVGSGQTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVGSGQTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
QY 153 VVAMNSGGGGTTPRVPGRVWALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGGPV 212
Db 128 VVAMNSGGGGTTPRVPGRVWALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGGPV 187
QY 213 VNLGQVVGNTAASDNFQSGGGTTPRVPGRVWALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGGPV 272
Db 188 VNLGQVVGNTAASDNFQSGGGTTPRVPGRVWALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGGPV 247
QY 273 GVVDNNGNGARVQVWGSAPAAISLIGTGVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDNNGNGARVQVWGSAPAAISLIGTGVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNWQTKSGGTRTGNVTIAEGPPA 355
Db 308 VNWQTKSGGTRTGNVTIAEGPPA 330
RESULT 15
AAE17566
ID AAE17566 standard; protein; 330 AA.
XX
XX AAE17566;
XX
XX 22-APR-2002 (first entry)
DT

XX Mycobacterium sp. MTB32A (Ra35FL) mature protein.
XX
XX Fusion protein; antigen; serological sensitivity; immune response;
XX tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
XX
XX Mycobacterium sp.
XX
XX Key Location/Qualifiers
XX Region 8..202
XX Misc-difference 182 /note= "Ra35 N-terminal peptide"
XX Misc-difference 183 /note= "Encoded by GAG"
XX Misc-difference 199..330 /note= "Encoded by GCG"
XX Region 199..330
XX Region /note= "Ra35 C-terminal peptide, Ra12"
XX
XX WO200198460-A2.
XX
XX 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-US019959.
XX
XX 20-JUN-2000; 2000US-00597796.
XX 01-FEB-2001; 2001US-0265737P.
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Alderson M;
XX WPI; 2002-147798/19.
XX N-PSDB; AAD28336.
XX
XX Composition comprising MTB39 antigen and MTB32A antigen from
XX Mycobacterium species, useful for eliciting immune response in a subject.
XX
XX Claim 70; Fig 6; 136pp; English.
XX
XX The present invention relates to fusion proteins containing at least two
XX Mycobacterium species antigens, nucleotides encoding them and
XX compositions comprising such fusion proteins. The present invention
XX particularly relates to nucleic acids encoding fusion proteins that
XX include two or more individual M. tuberculosis antigens which increase
XX the serological sensitivity of sera from individuals infected with
XX tuberculosis and methods for their use in diagnosis, prevention and
XX treatment of tuberculosis infection. Sequences of the invention are
XX useful for eliciting an immune response in a mammal, e.g., human,
XX immunised with BCG. They are useful in the diagnosis, treatment and
XX prevention of Mycobacterium infection. The fusion proteins and the
XX polynucleotides are useful as diagnostic tools in patients infected with
XX Mycobacterium, in vitro and in vivo assays for detecting humoral
XX antibodies or cell-mediated immunity against M. tuberculosis, for the
XX diagnosis of an infection or monitoring of disease progression, as
XX immunogens to generate or elicit a protective immune response in a
XX patient and for raising anti-M. tuberculosis antibodies in a non-human
XX animal. Sequences of the invention are also used as vaccines. MTB32A
XX fusion proteins of the invention are useful as in vivo diagnostic agents
XX for intradermal skin test. The present sequence is Mycobacterium species
XX MTB32A (Ra32FL) mature protein
XX
XX Sequence 330 AA;
XX

Query Match 90.7%; Score 1634; DB 5; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.1e-113;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNAVGAGTGIVDPNGVILT 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNAVGAGTGIVDPNGVILT 67
QY 93 NNHVIAGATDINAFSVGSGQTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152

Db 68 NNHVIAGATDINAFVSGSQTYGVDDVGYDRDQDVAVLQLRGAGGLPSAAIGCGVAVGEP 127
QY 153 VVAMGNSGGQGTTPRAVPRVVALGQTVQASDSLGTGABETLNGLIOPDAAIQGDSGGPV 212
Db 128 VVAMGNSGGQGTTPRAVPRVVALGQTVQASDSLGTGABETLNGLIQDAAIQGDSGGPV 187
QY 213 VNGLQGVVGMNTAASDNFQLSQGGQGFAPICQAMAIAGQIRSGGSGPTVHIGPTAFGL 272
Db 188 VNGLQGVVGMNTAASDNFQLSQGGQGFAPICQAMAIAGQIRSGGSGPTVHIGPTAFGL 247
QY 273 GVDDNNGNGARVORVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVDDNNGNGARVORVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
Db 308 VTWQTKSGGTRTGNVTLAEGPPA 330

Search completed: June 22, 2004, 17:15:23
Job time : 44.583 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 12.6725 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802

Sequence: 1 MNSRRRLRLWSNLLSVLAA.....QTKSGTGTGNTVLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PGTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1802	100.0	355	3	US-08-818-112-79
2	1802	100.0	355	4	US-08-818-111-80
3	1802	100.0	355	4	US-09-056-556-79
4	1802	100.0	355	4	US-09-072-596-80
5	1802	100.0	355	4	US-09-072-967-79
6	983	54.6	596	4	US-09-287-849-26
7	983	54.6	729	4	US-09-223-040-2
8	983	54.6	729	4	US-09-287-849-2
9	676	37.5	132	3	US-08-818-112-66
10	676	37.5	132	4	US-08-818-111-67
11	676	37.5	132	4	US-09-056-556-66
12	676	37.5	132	4	US-09-072-596-67
13	676	37.5	132	4	US-09-072-967-66
14	676	37.5	132	4	US-09-636-215-819
15	676	37.5	132	4	US-09-636-215-848
16	676	37.5	132	4	US-09-685-166A-819
17	676	37.5	132	4	US-09-685-166A-848
18	671	37.2	231	4	US-09-287-849-28
19	628	34.9	224	4	US-09-636-215-825
20	628	34.9	224	4	US-09-685-166A-825
21	628	34.9	267	4	US-09-643-597-352
22	628	34.9	267	4	US-09-606-421B-352
23	628	34.9	273	4	US-09-736-457-1864
24	628	34.9	299	4	US-09-643-597-354
25	628	34.9	299	4	US-09-606-421B-354
26	628	34.9	304	4	US-09-636-215-835
27	628	34.9	304	4	US-09-685-166A-835

28 628 34.9 314 4 US-09-736-457-1863 Sequence 1863, App
29 628 34.9 400 4 US-09-636-215-852 Sequence 852, App
30 628 34.9 400 4 US-09-685-166A-852 Sequence 852, App
31 628 34.9 487 4 US-09-620-412C-349 Sequence 349, App
32 628 34.9 487 4 US-09-598-419-349 Sequence 349, App
33 628 34.9 518 4 US-09-620-412C-333 Sequence 333, App
34 628 34.9 518 4 US-09-598-419-333 Sequence 333, App
35 628 34.9 525 4 US-09-556-877-196 Sequence 196, App
36 628 34.9 525 4 US-09-620-412C-196 Sequence 196, App
37 628 34.9 525 4 US-09-598-419-196 Sequence 196, App
38 628 34.9 583 4 US-09-620-412C-353 Sequence 353, App
39 628 34.9 583 4 US-09-598-419-353 Sequence 353, App
40 628 34.9 585 4 US-09-620-412C-337 Sequence 337, App
41 628 34.9 585 4 US-09-598-419-337 Sequence 337, App
42 628 34.9 619 4 US-09-620-412C-309 Sequence 309, App
43 628 34.9 619 4 US-09-598-419-309 Sequence 309, App
44 628 34.9 631 4 US-09-620-412C-325 Sequence 325, App
45 628 34.9 631 4 US-09-598-419-325 Sequence 325, App

ALIGNMENTS

RESULT 1
US-08-818-112-79
; Sequence 79, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-79

Query Match 100.0%; Score 1802; DB 3; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAAPALSDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLATAPAAAPALSDRFPALPLDPSAMVAQVA 60
QY 61 PQVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVVG 120
DB 61 PQVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVVG 120
QY 121 YRTOQVAVLQIRGAGGLPSAAIGGVAVGEPVAMGNSGGGTPRAVGRVVALGQTV 180
DB 121 YRTOQVAVLQIRGAGGLPSAAIGGVAVGEPVAMGNSGGGTPRAVGRVVALGQTV 180
QY 181 QASDSLTGAETLNGLIQFDDAAIQPGDSGGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
DB 181 QASDSLTGAETLNGLIQFDDAAIQPGDSGGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
QY 241 IPIGQAMAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGAPAAASLGIST 300
DB 241 IPIGQAMAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGAPAAASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVLSVWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGPINSATAMADALNGHHPGDVLSVWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-08-818-111-80
; Sequence 80, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-80

Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAAPALSDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLATAPAAAPALSDRFPALPLDPSAMVAQVA 60
QY 61 PQVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVVG 120
DB 61 PQVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVVG 120
QY 121 YRTOQVAVLQIRGAGGLPSAAIGGVAVGEPVAMGNSGGGTPRAVGRVVALGQTV 180
DB 121 YRTOQVAVLQIRGAGGLPSAAIGGVAVGEPVAMGNSGGGTPRAVGRVVALGQTV 180
QY 181 QASDSLTGAETLNGLIQFDDAAIQPGDSGGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
DB 181 QASDSLTGAETLNGLIQFDDAAIQPGDSGGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
QY 241 IPIGQAMAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGAPAAASLGIST 300
DB 241 IPIGQAMAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGAPAAASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVLSVWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGPINSATAMADALNGHHPGDVLSVWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-09-056-556-79
; Sequence 79, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79

Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAAPALSDRFPALPLDPSAMVAQVA 60

Db 1 MSNRRSLRWSLLSVLAAGVGLGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNTINTKLYNNVAGAGTGVIVDPNGVVLNNHVIAGATDINAFSGSGGTGYVDVVG 120
Db 61 PQVNTINTKLYNNVAGAGTGVIVDPNGVVLNNHVIAGATDINAFSGSGGTGYVDVVG 120
QY 121 YDRTQDVAVLQURGAGLPSAAIGGVAVGEPVAVMNGSGGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQURGAGLPSAAIGGVAVGEPVAVMNGSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCABEETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
Db 181 QASDSLTCABEETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
QY 241 IPGQAMATAGQIRSGGSGPTVHIGTAFGLGVVDNNGNGARVQVVGSAFAASLGIST 300
Db 241 IPGQAMATAGQIRSGGSGPTVHIGTAFGLGVVDNNGNGARVQVVGSAFAASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVSVNWQTKSGGTGTGNVTLAEGPPA 355
Db 301 GDVITAVDGPINSATAMADALNGHHPGDVSVNWQTKSGGTGTGNVTLAEGPPA 355

RESULT 4
US-09-072-596-80
; Sequence 80, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-80

Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNRRSLRWSLLSVLAAGVGLGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MSNRRSLRWSLLSVLAAGVGLGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNTINTKLYNNVAGAGTGVIVDPNGVVLNNHVIAGATDINAFSGSGGTGYVDVVG 120
Db 61 PQVNTINTKLYNNVAGAGTGVIVDPNGVVLNNHVIAGATDINAFSGSGGTGYVDVVG 120
QY 121 YDRTQDVAVLQURGAGLPSAAIGGVAVGEPVAVMNGSGGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQURGAGLPSAAIGGVAVGEPVAVMNGSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCABEETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
Db 181 QASDSLTCABEETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
QY 241 IPGQAMATAGQIRSGGSGPTVHIGTAFGLGVVDNNGNGARVQVVGSAFAASLGIST 300
Db 241 IPGQAMATAGQIRSGGSGPTVHIGTAFGLGVVDNNGNGARVQVVGSAFAASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVSVNWQTKSGGTGTGNVTLAEGPPA 355
Db 301 GDVITAVDGPINSATAMADALNGHHPGDVSVNWQTKSGGTGTGNVTLAEGPPA 355

RESULT 5
US-09-072-967-79
; Sequence 79, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-79

```
Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151; Indels 0; Gaps 0;
Matches 355; Conservative 0; Mismatches 0;

QY 1 MSNRRSLRWSLLSLVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MSNRRSLRWSLLSLVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60

QY 61 PQVNNINTKLGNNAVAGAGTGIVIDPQVLTNNHVIAGATDINAFSVSGQTYGVDDVVG 120
Db 61 PQVNNINTKLGNNAVAGAGTGIVIDPQVLTNNHVIAGATDINAFSVSGQTYGVDDVVG 120

QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPEVPMNGSGGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPEVPMNGSGGGGTTPRAVGRVVALGQTV 180

QY 181 QASDSLTGAEETLNGLIQFDAIAIQFSGSGGPPVNVNGLQGVVGMNTAASDNFQLSQGGGPPA 240
Db 181 QASDSLTGAEETLNGLIQFDAIAIQFSGSGGPPVNVNGLQGVVGMNTAASDNFQLSQGGGPPA 240

QY 241 IPIQMAIAIQIRSGSGSPVTHIGPTAFILGLGVDDNNGNGARVQRVVGSAAPASLGIST 300
Db 241 IPIQMAIAIQIRSGSGSPVTHIGPTAFILGLGVDDNNGNGARVQRVVGSAAPASLGIST 300

QY 301 GDVITAVDGAPINSATAMADALNGHPGCDVISVNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAPINSATAMADALNGHPGCDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 6
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIORITY FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 54.6%; Score 983; DB 4; Length 596;
Best Local Similarity 97.5%; Pred. No. 1.1e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQVPOVNNITKLGNNAVAGAGTGIVIDPN 87
Db 397 AAGDIAPPALSDQRFADFPALPLDPSAMVAQVPOVNNITKLGNNAVAGAGTGIVIDPN 456

RESULT 7
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

Query Match 54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred. No. 1.1e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQVPOVNNITKLGNNAVAGAGTGIVIDPN 87
Db 530 AAGDIAPPALSDQRFADFPALPLDPSAMVAQVPOVNNITKLGNNAVAGAGTGIVIDPN 589

QY 88 GVLTTNNHVIAGATDINAFSVSGQTYGVDDVYDRTQDVAVLQLRGAGGLPSAAIGGV 147
Db 590 GVLTTNNHVIAGATDINAFSVSGQTYGVDDVYDRTQDVAVLQLRGAGGLPSAAIGGV 649

QY 148 AVGEPVPMNGSGGGGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAIAIQPD 207
Db 650 AVGEPVPMNGSGGGGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAIAIQPD 709

QY 208 SGGPVVNGLGQVVGMMNTAAS 227
Db 710 SGGPVVNGLGQVVGMMNTAAS 729

RESULT 8
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIORITY FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 54.6%; Score 983; DB 4; Length 596;
Best Local Similarity 97.5%; Pred. No. 1.1e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQVPOVNNITKLGNNAVAGAGTGIVIDPN 87
Db 397 AAGDIAPPALSDQRFADFPALPLDPSAMVAQVPOVNNITKLGNNAVAGAGTGIVIDPN 456
```

```

; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match          54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred.No. 1.5e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQVAPQVNNINTKLGYNNVAGAGTGIVIDPN 87
Db 530 AAGDIAPPALSDQRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAGAGTGIVIDPN 589

QY 88 GVLTNNHVIAGATDINAFSVGSQGYGVYDVVGYDRTQDVAVLQLRGAGGLPSAATGGGV 147
Db 590 GVLTNNHVIAGATDINAFSVGSQGYGVYDVVGYDRTQDVAVLQLRGAGGLPSAATGGGV 649

QY 148 AVGEPVVMGNSGGQGTFRVGRVVALGQTVQASDSLTGABETLNGLIQFDDAATQPGD 207
Db 650 AVGEPVVMGNSGGQGTFRVGRVVALGQTVQASDSLTGABETLNGLIQFDDAATQPGD 709

QY 208 SGGPVVNGLGQVVGMMTAA 227
Db 710 SGGPVVNGLGQVVGMMTAA 729

RESULT 9
US-08-818-112-66
; Sequence 66, Application US/08818112
; Patent No. 6290569
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997

Query Match          37.5%; Score 676; DB 3; Length 132;
Best Local Similarity 100.0%; Pred.No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFOLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGAR 283
Db 1 TAASDNFOLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGAR 60

QY 284 VQVVVGSAPAASLGTSGDVITAVDGPINSATAMADALNGHHPGCDVTSVNWQTKSGGTR 343
Db 61 VQVVVGSAPAASLGTSGDVITAVDGPINSATAMADALNGHHPGCDVTSVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 10
US-08-818-111-67
; Sequence 67, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
```

```
;
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-67

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60

QY 284 VQRYVGSAPAAASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVVISVNWQTKSGGTR 343
Db 61 VQRYVGSAPAAASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 11
US-09-056-556-66
; Sequence 66, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-66

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60

QY 284 VQRYVGSAPAAASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVVISVNWQTKSGGTR 343
Db 61 VQRYVGSAPAAASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

TREATM

;
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-67

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60

QY 284 VQRYVGSAPAAASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVVISVNWQTKSGGTR 343
Db 61 VQRYVGSAPAAASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 12
US-09-072-596-67
; Sequence 67, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Meto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-67

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60

QY 284 VQRYVGSAPAAASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVVISVNWQTKSGGTR 343
Db 61 VQRYVGSAPAAASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 13
US-09-072-967-66
```

; Sequence 66, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-66

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFOLSGGCGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 283
Db 1 TAASDNFOLSGGCGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 60

QY 284 VQRVVGSAAPASLGITGDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
Db 61 VQRVVGSAAPASLGITGDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 14
US-09-636-215-819
; Sequence 819, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui

; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stoik, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 819
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-636-215-819

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFOLSGGCGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 283
Db 1 TAASDNFOLSGGCGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 60

QY 284 VQRVVGSAAPASLGITGDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
Db 61 VQRVVGSAAPASLGITGDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 15
US-09-636-215-848
; Sequence 848, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stoik, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 848
; LENGTH: 132

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-636-215-848

Query Match      37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred.No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 TAASDNFQLSQGGGFAIPIGQAWAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGAR 283
    |||||
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGAR 60
    |||||

Qy 284 VQRVVGSAAPASLGISTGVDVITAVDGPAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
    |||||
Db 61 VQRVVGSAAPASLGISTGVDVITAVDGPAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120
    |||||

Qy 344 TGNVTLAEGGPPA 355
    |||||
Db 121 TGNVTLAEGGPPA 132
    |||||

```

Search completed: June 22, 2004, 17:27:28
Job time : 13.6725 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 32.8799 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802
Sequence: 1 MSNRRSLRWLSVLA.....QPKSGGTGNTVLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 28213646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PC7_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/CTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1802	100.0	355	12	US-09-886-349A-2
2	1802	100.0	355	14	US-10-193-002-80
3	1802	100.0	355	14	US-10-084-843-79
4	1802	100.0	355	14	US-10-098-732A-2
5	1792	99.4	355	9	US-09-712-363-161
6	1634	90.7	330	12	US-09-886-349A-4
7	1634	90.7	330	14	US-10-098-732A-4
8	1634	90.7	330	15	US-10-369-983-19
9	1631	90.5	330	12	US-09-886-349A-6
10	1631	90.5	330	14	US-10-098-732A-6
11	1631	90.5	330	15	US-10-369-983-20
12	1631	90.5	723	15	US-10-369-983-2
13	1631	90.5	1010	15	US-10-369-983-4
14	999.5	55.5	1016	15	US-10-369-983-18
15	992	55.0	1154	15	US-10-369-983-16

16	990.5	55.0	825	15	US-10-369-983-14	Sequence 14, Appl
17	989.5	54.9	813	15	US-10-369-983-15	Sequence 15, Appl
18	988	54.8	1022	15	US-10-369-983-17	Sequence 17, Appl
19	985	54.7	195	12	US-09-886-349A-8	Sequence 8, Appl
20	985	54.7	195	14	US-10-098-732A-8	Sequence 8, Appl
21	983.5	54.6	875	15	US-10-369-983-13	Sequence 13, Appl
22	983	54.6	596	9	US-09-287-849-26	Sequence 26, Appl
23	983	54.6	596	12	US-09-886-349A-20	Sequence 20, Appl
24	983	54.6	596	14	US-10-359-460-26	Sequence 26, Appl
25	983	54.6	596	14	US-10-098-732A-20	Sequence 20, Appl
26	983	54.6	729	9	US-09-287-849-2	Sequence 2, Appl
27	983	54.6	729	12	US-09-886-349A-16	Sequence 16, Appl
28	983	54.6	729	14	US-10-359-460-2	Sequence 2, Appl
29	983	54.6	729	14	US-10-098-732A-16	Sequence 16, Appl
30	983	54.6	729	15	US-10-369-983-21	Sequence 21, Appl
31	983	54.6	729	15	US-10-359-459-2	Sequence 2, Appl
32	983	54.6	930	14	US-10-098-732A-65	Sequence 65, Appl
33	983	54.6	930	15	US-10-369-983-12	Sequence 12, Appl
34	980	54.4	729	12	US-09-886-349A-18	Sequence 18, Appl
35	980	54.4	729	14	US-10-098-732A-18	Sequence 18, Appl
36	980	54.4	729	15	US-10-369-983-22	Sequence 22, Appl
37	676	37.5	132	9	US-09-759-143-819	Sequence 819, App
38	676	37.5	132	9	US-09-759-143-848	Sequence 848, App
39	676	37.5	132	9	US-09-780-669-819	Sequence 819, App
40	676	37.5	132	9	US-09-780-669-848	Sequence 848, App
41	676	37.5	132	9	US-09-822-827-819	Sequence 819, App
42	676	37.5	132	9	US-09-822-827-848	Sequence 848, App
43	676	37.5	132	9	US-09-895-793-819	Sequence 819, App
44	676	37.5	132	9	US-09-895-793-848	Sequence 848, App
45	676	37.5	132	9	US-09-895-814-819	Sequence 819, App

ALIGNMENTS

RESULT 1

US-09-886-349A-2
; Sequence 2, Application US/09886349A
; Publication No. US200040086523A1
; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35FL)
US-09-886-349A-2

Query Match 100.0%; Score 1802; DB 12; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSNRRSLRWLSVLA	AVGLGLATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA	60
Db	1	MSNRRSLRWLSVLA	AVGLGLATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA	60
QY	61	POVNVNITKLYNNA	VAGTGTIVDPNGVLTNNHVIAGTDINAFSGSQTTGVDDVWG	120
Db	61	POVNVNITKLYNNA	VAGTGTIVDPNGVLTNNHVIAGTDINAFSGSQTTGVDDVWG	120

QY 121 YDRTDQVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGGTTPRAVGRVVALGQTV 180
 Db 121 YDRTDQVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTCABETLNLQIFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGGFA 240
 Db 181 QASDSLTCABETLNLQIFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGGFA 240
 QY 241 IPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPAASLGIST 300
 Db 241 IPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPAASLGIST 300
 QY 301 GDVITAVDGAIPINSATAMADALNGHHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGAIPINSATAMADALNGHHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-10-193-002-80
 ; Sequence 80, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonia
 ; Houghton, Raymond
 ; Vedvick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 TUBERCULOSIS

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/193.002
 FILING DATE: 10-Jul-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072.596
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 80:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 80:

US-10-193-002-80

Query Match

Best Local Similarity 100.0%; Score 1802; DB 14; Length 355;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNRRSLRWLSWLLSVLAAYGLGLATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA 60
 Db 1 MSNRRSLRWLSWLLSVLAAYGLGLATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNINIKLYNNNAVAGAGTGIVIDPNGVLTNNHVIAGATDINAFVSGSQTYGVVDVVG 120
 Db 61 PQVNINIKLYNNNAVAGAGTGIVIDPNGVLTNNHVIAGATDINAFVSGSQTYGVVDVVG 120
 QY 121 YDRTDQVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGGTTPRAVGRVVALGQTV 180
 Db 121 YDRTDQVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTCABETLNLQIFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGGFA 240
 Db 181 QASDSLTCABETLNLQIFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGGFA 240
 QY 241 IPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPAASLGIST 300
 Db 241 IPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPAASLGIST 300
 QY 301 GDVITAVDGAIPINSATAMADALNGHHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGAIPINSATAMADALNGHHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-10-084-843-79
 ; Sequence 79, Application US/10084843
 ; Publication No. US20030143243A1
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonio
 ; Houghton, Raymond
 ; Vedvick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,843
 FILING DATE: 25-Feb-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: amino acid


```
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-084-843-79

Query Match      100.0%; Score 1802; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
QY 61 PQVNTNKLGYNNVAGAGTGVIDPVGVLVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
DB 61 PQVNTNKLGYNNVAGAGTGVIDPVGVLVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
QY 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCABETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFOLSGGGQGFA 240
DB 181 QASDSLTCABETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFOLSGGGQGFA 240
QY 241 IPTGQAMAIAGQIRSGGSPTHVIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
DB 241 IPTGQAMAIAGQIRSGGSPTHVIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4
US-10-098-732A-2
; Sequence 2, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; CURRENT APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35FL)
US-10-098-732A-2

Query Match      100.0%; Score 1802; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
QY 61 PQVNTNKLGYNNVAGAGTGVIDPVGVLVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
DB 61 PQVNTNKLGYNNVAGAGTGVIDPVGVLVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
QY 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
```

```
DB 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCABETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFOLSGGGQGFA 240
DB 181 QASDSLTCABETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFOLSGGGQGFA 240
QY 241 IPTGQAMAIAGQIRSGGSPTHVIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
DB 241 IPTGQAMAIAGQIRSGGSPTHVIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 5
US-09-712-363-161
; Sequence 161, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Edward M.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-161

Query Match      99.4%; Score 1792; DB 9; Length 355;
Best Local Similarity 99.4%; Pred. No. 9.6e-134;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQV 60
QY 61 PQVNTNKLGYNNVAGAGTGVIDPVGVLVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
DB 61 PQVNTNKLGYNNVAGAGTGVIDPVGVLVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
QY 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCABETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFOLSGGGQGFA 240
```

Db 181 QASDLSLGAETLNLGIQFDAAIQPGDSGGPVVNGLVGVVGMNTAASDNFQLSQGGGPPA 240
 Qy 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVRVGSAPAAISLGIST 300
 Db 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVRVGSAPAAISLGIST 300
 Qy 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355
 Db 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355

RESULT 6

US-09-886-349A-4
 ; Sequence 4, Application US/09886349A
 ; Publication No. US20040086523A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009070US
 ; CURRENT APPLICATION NUMBER: US/09/886,349A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: US 09/597,796
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: US 60/265,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 330
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; OTHER INFORMATION: MTB32A (Ra35 mature)
 US-09-886-349A-4

Query Match 90.7%; Score 1634; DB 12; Length 330;
 Best Local Similarity 99.4%; Pred. No. 2.9e-121;
 Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFALPLDPSAMVAQVAVQVNVNINIKLGYNNVAVGAGTGIVDPNGVLT 92
 Db 8 APPALSQDRFADFALPLDPSAMVAQVAVQVNVNINIKLGYNNVAVGAGTGIVDPNGVLT 67
 Qy 93 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 152
 Db 68 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 127
 Qy 153 VVMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNLGIQFDAAIQPGDSGGPV 212
 Db 128 VVMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNLGIQFDAAIQPGDSGGPV 187
 Qy 213 VNLGQVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 272
 Db 188 VNLGQVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 247
 Qy 273 GVVDNNGGARVQVRVGSAPAAISLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 332
 Db 248 GVVDNNGGARVQVRVGSAPAAISLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 307
 Qy 333 VNWQTKSGGTRTGNVTLAGGPPA 355
 Db 308 VTWQTKSGGTRTGNVTLAGGPPA 330

RESULT 7

US-10-098-732A-4
 ; Sequence 4, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir

; APPLICANT: Brannon, Mark
 ; APPLICANT: Guderian, Jeffrey
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; TITLE OF INVENTION: Leishmania Antigen
 ; FILE REFERENCE: 014058-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 330
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Ra35 mature
 US-10-098-732A-4

Query Match 90.7%; Score 1634; DB 14; Length 330;
 Best Local Similarity 99.4%; Pred. No. 2.9e-121;
 Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFALPLDPSAMVAQVAVQVNVNINIKLGYNNVAVGAGTGIVDPNGVLT 92
 Db 8 APPALSQDRFADFALPLDPSAMVAQVAVQVNVNINIKLGYNNVAVGAGTGIVDPNGVLT 67
 Qy 93 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 152
 Db 68 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 127
 Qy 153 VVMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNLGIQFDAAIQPGDSGGPV 212
 Db 128 VVMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNLGIQFDAAIQPGDSGGPV 187
 Qy 213 VNLGQVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 272
 Db 188 VNLGQVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 247
 Qy 273 GVVDNNGGARVQVRVGSAPAAISLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 332
 Db 248 GVVDNNGGARVQVRVGSAPAAISLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 307
 Qy 333 VNWQTKSGGTRTGNVTLAGGPPA 355
 Db 308 VTWQTKSGGTRTGNVTLAGGPPA 330

RESULT 8

US-10-369-983-19
 ; Sequence 19, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 19
 ; LENGTH: 330
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: wild-type
 ; OTHER INFORMATION: mature MTB32A (Ra35)

US-10-369-983-19

Query Match 90.7%; Score 1634; DB 15; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVVIDPNGVVL 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVVIDPNGVVL 67
QY 93 NNHVIAGATDINAFSGSGQTYGVVDYDRDQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSGSGQTYGVVDYDRDQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 127
QY 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
QY 213 VNLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 272
DB 188 VNLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 247
QY 273 GVVDNNGGARVQVRVVGSAAPASLSIGISTGDIVITAVDGAIPNSATAMADALNGHHFPGDVIS 332
DB 248 GVVDNNGGARVQVRVVGSAAPASLSIGISTGDIVITAVDGAIPNSATAMADALNGHHFPGDVIS 307
QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 9

US-09-886-349A-6

; Sequence 6, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA
US-09-886-349A-6

Query Match 90.5%; Score 1631; DB 12; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVVIDPNGVVL 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVVIDPNGVVL 67
QY 93 NNHVIAGATDINAFSGSGQTYGVVDYDRDQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSGSGQTYGVVDYDRDQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 127
QY 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187

QY 213 VNLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 272
DB 188 VNLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 247
QY 273 GVVDNNGGARVQVRVVGSAAPASLSIGISTGDIVITAVDGAIPNSATAMADALNGHHFPGDVIS 332
DB 248 GVVDNNGGARVQVRVVGSAAPASLSIGISTGDIVITAVDGAIPNSATAMADALNGHHFPGDVIS 307
QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 10

US-10-098-732A-6
; Sequence 6, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA
US-10-098-732A-6

Query Match 90.5%; Score 1631; DB 14; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVVIDPNGVVL 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVVIDPNGVVL 67
QY 93 NNHVIAGATDINAFSGSGQTYGVVDYDRDQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSGSGQTYGVVDYDRDQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 127
QY 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
QY 213 VNLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 272
DB 188 VNLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 247
QY 273 GVVDNNGGARVQVRVVGSAAPASLSIGISTGDIVITAVDGAIPNSATAMADALNGHHFPGDVIS 332
DB 248 GVVDNNGGARVQVRVVGSAAPASLSIGISTGDIVITAVDGAIPNSATAMADALNGHHFPGDVIS 307
QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 11

US-10-369-983-20
; Sequence 20, Application US/10369983
; Publication No. US20030235593A1

; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 20
 ; LENGTH: 330
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: mutated
 ; OTHER INFORMATION: MTB32MutSA (Ra35 mutSA)
 US-10-369-983-20

Query Match 90.5%; Score 1631; DB 15; Length 330;
 Best Local Similarity 99.1%; Pred. No. 4.9e-121;
 Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPNGVWLT 92
 DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPNGVWLT 67
 QY 93 NNHVIAGATDINAFSVSGGQTYGVVDVVGVDYRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
 DB 68 NNHVIAGATDINAFSVSGGQTYGVVDVVGVDYRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
 QY 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQDGDGGPV 212
 DB 128 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQDGDGGPV 187
 QY 213 VNLGQVVGVMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGL 272
 DB 188 VNLGQVVGVMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGL 247
 QY 273 GVDNNGNGARVORVVGSAAPASISLGISTGDIVITAVDGAIPNSATAMADALNGHHPGDVIS 332
 DB 248 GVDNNGNGARVORVVGSAAPASISLGISTGDIVITAVDGAIPNSATAMADALNGHHPGDVIS 307
 QY 333 VNMQTKSGGTRTGNVTILAEPPA 355
 DB 308 VTWQTKSGGTRTGNVTILAEPPA 330

RESULT 12
 US-10-369-983-2
 ; Sequence 2, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 723
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: mutated
 ; OTHER INFORMATION: MTB32-MTB39F fusion protein (MTB32MutSA)
 US-10-369-983-2

Query Match 90.5%; Score 1631; DB 15; Length 723;
 Best Local Similarity 99.1%; Pred. No. 1.3e-120;
 Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPNGVWLT 92
 DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPNGVWLT 67
 QY 93 NNHVIAGATDINAFSVSGGQTYGVVDVVGVDYRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
 DB 68 NNHVIAGATDINAFSVSGGQTYGVVDVVGVDYRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
 QY 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQDGDGGPV 212
 DB 128 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQDGDGGPV 187
 QY 213 VNLGQVVGVMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGL 272
 DB 188 VNLGQVVGVMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGL 247
 QY 273 GVDNNGNGARVORVVGSAAPASISLGISTGDIVITAVDGAIPNSATAMADALNGHHPGDVIS 332
 DB 248 GVDNNGNGARVORVVGSAAPASISLGISTGDIVITAVDGAIPNSATAMADALNGHHPGDVIS 307
 QY 333 VNMQTKSGGTRTGNVTILAEPPA 355
 DB 308 VTWQTKSGGTRTGNVTILAEPPA 330

RESULT 13
 US-10-369-983-4
 ; Sequence 4, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 1010
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: MTB-102F fusion
 ; OTHER INFORMATION: protein
 US-10-369-983-4

Query Match 90.5%; Score 1631; DB 15; Length 1010;
 Best Local Similarity 99.1%; Pred. No. 2e-120;
 Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPNGVWLT 92
 DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPNGVWLT 67
 QY 93 NNHVIAGATDINAFSVSGGQTYGVVDVVGVDYRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
 DB 68 NNHVIAGATDINAFSVSGGQTYGVVDVVGVDYRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
 QY 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQDGDGGPV 212

Db 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPDAGGPV 187
QY 213 VNLGLGVGWNVAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLLQL 272
Db 188 VNLGLGVGWNVAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLLQL 247
QY 273 GVVDNNGGARVQVVGSAASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDNNGGARVQVVGSAASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNMOTKSGGTRGNVTLAEGPPA 355
Db 308 VTWTKSGGTRGNVTLAEGPPA 330

RESULT 14
US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 55.5%; Score 999.5; DB 15; Length 1016;
Best Local Similarity 84.1%; Pred. No. 2e-70;
Matches 207; Conservative 6; Mismatches 18; Indels 15; Caps 4;
QY 28 APAQAAPPALSDRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNNAVAGTGIVIDPN 87
Db 530 AAGDIAPPALSDRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNNAVAGTGIVIDPN 589
QY 88 GVLTNNHVIAGATDINAFSVGSGQTYGVDVDRDQDVAVLQLRGAGGLPSAAIIGGV 147
Db 590 GVLTNNHVIAGATDINAFSVGSGQTYGVDVDRDQDVAVLQLRGAGGLPSAAIIGGV 649
QY 148 AVGEPVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPD 207
Db 650 AVGEPVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPD 709
QY 208 SGGPVVNLGQVVGWNTAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPT 267
Db 710 SGGPVVNLGQVVGWNTAASGTF--SRPG----LPVEYLQVPSPSGMRDIXVQFSGGNN 763
QY 259 SPTVHI 264
Db 764 SPAVYL 769

RESULT 15
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-16
Query Match 55.0%; Score 992; DB 15; Length 1154;
Best Local Similarity 63.8%; Pred. No. 9.2e-70;
Matches 219; Conservative 17; Mismatches 59; Indels 48; Gaps 6;
QY 28 APAQAAPPALSDRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNNAVAGTGIVIDPN 87
Db 530 AAGDIAPPALSDRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNNAVAGTGIVIDPN 589
QY 88 GVLTNNHVIAGATDINAFSVGSGQTYGVDVDRDQDVAVLQLRGAGGLPSAAIIGGV 147
Db 590 GVLTNNHVIAGATDINAFSVGSGQTYGVDVDRDQDVAVLQLRGAGGLPSAAIIGGV 649
QY 148 AVGEPVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPD 207
Db 650 AVGEPVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPD 709
QY 208 SGGPVVNLGQVVGWNTAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPT 267
Db 710 SGGPVVNLGQVVGWNTAASGTFMDF-----GLLPP-----EVNNGRMYSGGPESMLAAA 760
QY 268 AFLGLGVVDNNGGARVQVVGSAASLSLTGSDVITAVDGPIN--NSATAMADALNGH 325
Db 761 AWDGV-----AELTSAAVSYGVSVSTLIVEPMMGPAAAAAATAATPY 803
QY 326 HPGDVISVNW-----QTKSGGTRGNVTLAEGPPA 355
Db 804 -----VGLAATAALAKETATQAAAAAFAFGTAFAMTVPPS 839
Search completed: June 22, 2004, 18:07:51
Job time : 34.8799 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 9.93246 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-2
Perfect score: 1802
Sequence: 1 MNSRRSLRWLSVLAA.....QTKSGTGTGNVTLAEGPPA 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR 78: *
1: Piri: *
2: Piri: *
3: Piri: *
4: Piri: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.4	355	2 F70983	probable serine pr
2	1306	72.5	361	2 S47170	hypothetical prote
3	1272.5	70.6	354	2 A87242	probable secreted
4	468.5	26.0	464	2 C70821	probable serine pr
5	460	25.5	382	2 H86930	probable secreted
6	460	25.5	452	2 T45448	probable serine pr
7	382	21.2	394	2 S74643	proteinase rhoA (E
8	372	20.6	407	2 AG2150	serine proteinase
9	355.5	19.7	452	2 S77538	serine proteinase
10	354	19.6	433	2 H97199	htra-like serine p
11	345	19.1	362	2 T35257	probable secreted
12	343	19.0	457	2 AG0433	proteinase (EC 3.4
13	341.5	19.0	514	2 A82591	periplasmic protei
14	335.5	18.6	429	2 AD1894	serine proteinase
15	331	18.4	416	2 AB2057	serine proteinase
16	331	18.4	441	2 B75357	probable periplasm
17	331	18.4	455	2 C91142	serine endoprotein
18	331	18.4	455	2 F85987	serine endoprotein
19	329.5	18.3	474	2 F83550	serine proteinase
20	329	18.3	398	2 B71284	probable periplasm
21	329	18.3	455	2 AB0909	serine protease (E
22	328	18.2	355	1 JG6052	trypsin-like prote
23	328	18.2	355	2 D91142	trypsin-like prote
24	327	18.1	352	2 B82307	trypsin-like prote
25	327	18.1	355	2 G85987	trypsin-like prote
26	326	18.1	455	2 JG6051	heat shock protein
27	322	17.9	475	1 S15337	protease DO precur
28	319	17.7	475	2 AC0528	proteinase DO (EC
29	317.5	17.6	474	2 S45229	

30 317.5 17.6 474 2 E85500 proteinase DO (EC
31 317.5 17.6 474 2 E90649 proteinase DO (EC
32 316.5 17.6 348 2 H96556 serine protease Do
33 315.5 17.5 474 1 I40059 htra-like protein
34 315.5 17.5 474 2 A13349 proteinase DO (EC
35 313.5 17.4 530 2 F87590 serine proteinase
36 312.5 17.3 356 2 SC0909 trypsin-like prote
37 311.5 17.3 416 2 S75445 proteinase rhoB (E
38 311 17.3 401 2 AD2451 serine proteinase
39 310 17.2 408 2 H86891 exported serine pr
40 308.5 17.1 481 2 AH0410 global stress requ
41 306 17.0 472 2 C87408 serine proteinase
42 306 17.0 513 2 AD3418 proteinase DO (EC
43 305 16.9 328 1 D69109 serine proteinase
44 305 16.9 513 2 I40060 serine proteinase
45 303 16.8 499 2 B81914 probable periplasm

ALIGNMENTS

RESULT 1

F70983
Probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: F70983
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Conor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Kajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whithead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70983
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Cross-references: 1-355 <COL>
A:Cross-references: GB:Z96071; GB:AL123456; NID:g3242254; PIDN:CA809453.1; PID:g2181957
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: pepA
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp:

Query Match 99.4%; Score 1792; DB 2; Length 355;
Best Local Similarity 99.4%; Pred. No. 7.5e-98;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MNSRRSLRWLSVLAAVGLGLATAPAAQAPALQSDRFPADFPALPLDPSAMVAQVA 60
Db 1 MNSRRSLRWLSVLAAVGLGLATAPAAQAPALQSDRFPADFPALPLDPSAMVAQVG 60
QY 61 PQVNVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
Db 61 PQVNVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY 121 YRTOQDVAVLQIRAGGLPSAAIGGVAVGEVWAMNSGGGGTTPRVPGRVWALQTV 180
Db 121 YRTOQDVAVLQIRAGGLPSAAIGGVAVGEVWAMNSGGGGTTPRVPGRVWALQTV 180
QY 181 QASDSLTGAETLNGLIQFDAAIQPDSGGPVYVGLGVVGMNTAASDNFQLSQGGQGA 240
Db 181 QASDSLTGAETLNGLIQFDAAIQPDSGGPVYVGLGVVGMNTAASDNFQLSQGGQGA 240
QY 241 IFIGQAMATAGIRSGGSPVTHIGTAPFLGLGVVDNNGARVORVVGSAASAIGIST 300
Db 241 IFIGQAMATAGIRSGGSPVTHIGTAPFLGLGVVDNNGARVORVVGSAASAIGIST 300
QY 301 GDVITAVDCAPINSATAMADALNGHHHPQDVISVNWOTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDCAPINSATAMADALNGHHHPQDVISVNWOTKSGGTRTGNVTLAEGPPA 355

```
RESULT 2
S47170
hypothetical protein 34K - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Dec-2002
C:Accession: S47170
R:Camaron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
submitted to the EMBL Data Library, June 1993
A:Description: Isolation and characterisation of a 34KDa protein of Mycobacterium paratu
A:Reference number: S47170
A:Accession: S47170
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <CAM>
A:Cross-references: EMBL:Z23092; NID:G505550; PIDN:CAA80638.1; PID:G505551
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
Query Match 72.5%; Score 1306; DB 2; Length 361;
Best Local Similarity 70.4%; Pred. No. 1.9e-69;
Matches 255; Conservative 41; Mismatches 58; Indels 8; Gaps 3;
QY 1 MSNS-RRRSRLWSWLLSVLAAGVGLG-----ATAPAAQAPALSDQRFADFPALPLDPS 53
Db 1 MSKSHHRSVWWSWLVGLTVVGLGSGVGLAPASAPSGUALDRFADRLPLDPS 60
QY 54 AMVAQVAPQVNTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGQT 113
Db 61 AMVGQVGPVNNIDTFKGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGQT 120
QY 114 YGVDMVGVDRTOQVAVLQRLGAGGLPSAATCGGVAVGEPVAMGNSGGCGGTFRVPGRV 173
Db 121 YAVDMVGVDRTOQVAVLQRLGAGGLPSAATCGGVAVGEPVAMGNSGGCGGTFRVPGRV 180
QY 174 VALGQTVQASDSLTAAGETLGLIQFDAAIQPDGSGGPPVNGLVGVVGMNTAASDNFQLS 233
Db 181 VALNQSVSATDTLTAAGENLGLIQADAPIKPDGSGGPPVNGLVGVVGMNTAATDYSKMS 240
QY 234 QGGGGAIPQQAIAAGQIRSGGSGTTHIGPTAFGLGVNDNGNGARVQRVWSGAPA 293
Db 241 -GGGGAIPQIRAWAVANQIRSGAGSTVHIGPTAFGLGVNDNGNGARVQRVWNTGPA 299
QY 294 ASLGISTGSDVITAVDGAIPNSATAMADALNGHHPGDVVISVWOTKSGTGTGNTVLAEGP 353
Db 300 AAGIAPGSDVITGVTPINGATSMTEVLVPHFGDTIAVHFSVDGGERTANITLAEGP 359
QY 354 PA 355
Db 360 PA 361
RESULT 3
A87242
Probable secreted serine proteinase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C:Accession: A87242
R:Coile, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
cam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: A87242
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <STO>
A:Cross-references: GB:AL450380; NID:G13093863; PIDN:CAC32191.1; GSPDB:GN00147
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
Query Match 70.6%; Score 1272.5; DB 2; Length 354;
```

```
Best Local Similarity 70.1%; Pred. No. 1.7e-67;
Matches 251; Conservative 42; Mismatches 58; Indels 7; Gaps 3;
QY 1 MSNSRRRSRLWSWLLSVLAAGVGLGATAPAAQ---APPALSDQRFADFPALPLDPSAMVA 57
Db 1 MSRQPHRSRLWSWLLSVLAAGVGLGSLAVVPGSATPSGSPSTLALDRFSNRPLPLNPAAM-- 58
QY 58 QVAPQVNTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTQYVD 117
Db 59 -VAPQVNTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTQYVD 117
QY 118 VVGYDRTOQVAVLQRLGAGGLPSAATCGGVAVGEPVAMGNSGGCGGTFRVPGRVWALG 177
Db 118 VVGYDRTOQVAVLQRLGAGGLPSAATCGGVAVGEPVAMGNSGGCGGTFRVPGRVWALN 177
QY 178 QTVQASDSLTAAGETLGLIQFDAAIQPDGSGGPPVNGLVGVVGMNTAASDNFQLSQGG 237
Db 178 QTVQASEPLTGAQETLGLIQVDAPIKPDGSGGPPVNSRGQVVGMTAATDNYKM-LGGQ 236
QY 238 GFAPITGAQALAGQIRSGGSGPTVHIGPTAFGLGVNDNGNGARVQRVWSGAPASLG 297
Db 237 GFAPITGAQMEVVGAIKSGAGSTVHIGPTAFGLGVNDNGNGARVQRVWSGAPASAG 296
QY 298 ISTGDTVAVDGAIPNSATAMADALNGHHPGDVVISVWOTKSGTGTGNTVLAEGPPA 355
Db 297 ISVGDITSDGVPISEATAMTNLVPHPGPTVAVNYSAGGGDLTANVLAEGPPA 354
RESULT 4
C70821
probable serine proteinase Rv0983 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: C70821
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70821
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-464 <COL>
A:Cross-references: GB:AL021999; GB:AL123456; NID:G3261536; PIDN:CAA17582.1; PID:e125391(
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv0983
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
Query Match 26.0%; Score 468.5; DB 2; Length 464;
Best Local Similarity 36.0%; Pred. No. 2.1e-20;
Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;
QY 19 AAVGLGLATAQAAPPALSDQRFADFPALPLDPSA---MVAQVAPQVNTKLYNNYNA 75
Db 120 AASLVGNRPAAGSGGPPVAAASAPSPIPANMPGSGVEQVAAKVVPSVWMLTDLGRQSE 179
QY 76 VGAGTGIVDPNGVLTNNHVIAGAT-----DINAFSVSGGTQYVDVGVYDRTOQV 127
Db 180 --EGSGIILSAEGLILTNHVIATAAKPPLGSPPPKTTVTFSDGRTAPFTVVGADPTSDI 237
QY 128 AVLQRLGAGGLPSAATCGG--VAVGEPVAMGNSGGCGGTFRVPGRVWALGQTVQASDS 185
Db 238 AVVRVQGVSGTLPISLGSSDLRVGQVLAIGSPLEGT--VTTGIVSALNRPVSTIGE 295
QY 186 LTGAETLGLIQFDAAIQPDGSGGPPVNGLVGVVGMNTA----ASDNFQLSQG--CQGF 239
Db 296 -AGNQNTVLDAIQDAAINFGNSGALVNNNAQLVGVNSAIATLIGADSADQAQSGISGLGF 354
QY 240 AIPITGAQALAGQIRSGGSGPTVHIGPTAFGLGVV--DNNNGNGARVQRVWSGAPASLGI 298
```


Db 355 AIPVDOAKRIADELISTGKA-----SHASLGVQVNTNDKDLTGAKIVEVWAGGAANAGV 408

QY 299 STGSDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAE 351

Db 409 PKGVVTKVDDREINSADALVAARSKAPGATVLTFODPGSGSRTVQVTLGK 461

RESULT 5

H86930

Probable secreted serine proteinase [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002

C:Accession: H86930

R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A: Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.

A: Title: Massive gene decay in the leprosy bacillus.

A: Reference number: A86909; PMID:21128732; PMID:11234002

A: Accession: H86930

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-382 <STO>

A: Cross-references: GB:AL450380; NID:gl3092536; PIDN:CAC29694.1; GSPDB:GN00147

C: Genetics:

A: Gene: ML0176

C: Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypsin

Query Match 25.5%; Score 460; DB 2; Length 382;

Best Local Similarity 36.4%; Pred. No. 5.3e-20;

Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

Db 22 GLGLATAPAAQAPALSDQDFADFPALPLDPSAMVAQA-----PQVNVNTKLGYNNAV 77

50 GAGPVTGPAASVPAANM-----PSGVEQVAVKVPVSVVMTDLGRQSE-- 94

QY 78 AGTGVIDPNGVLTNNHVIAGA-----TDINAFVSGSGQTVGVVGVYDR 123

95 EGSVILSADGLITNNHVAVAAPGPGGGLSPKTTVTFF---DGRFASFTVVGADP 151

QY 124 TDVAVQLRQAGGLPSAAICGG--VAVGEPVAVMGNSGGGTPPRAVGRVVALGQTVQ 181

152 TSDIAVVRVQISGLTPTITMGSSADLRVGPVAVGSPGLAGT--VTSGIVSALNRVPS 209

QY 182 ASDSLTGAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGMTAA-----SDNFQLSQG 235

210 TTGE-SGNQNTVLDALQTDAAINFGNSGGALVNMGGQLVGNVSAIATLGADSGDAQSGSI 268

QY 236 GQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGN-GARVQRVVGSAAPAA 294

269 GLGFAIPVDQAKRIADELISTG--KATH-----ASLGVQVATDKGTPGAKVMDVWAGGAA 322

QY 295 SLGISTGDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAE 351

323 NAAVFKGVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDQSGSSRTVQVTLGK 379

Db 323 NAAVFKGVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDQSGSSRTVQVTLGK 379

RESULT 6

T45448

Probable serine proteinase (EC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Dec-2002

C:Accession: T45448

R: James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, February 1998

A: Reference number: 222967

A: Accession: T45448

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA

A: Residues: 1-452 <JAM>

A: Cross-references: EMBL:AL035500; PIDN:CAB36690.1

A: Experimental source: cosmid L373

C: Genetics:

A: Note: MLCB373.28

C: Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypsin

C: Keywords: hydrolase; serine proteinase

F: 182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 25.5%; Score 460; DB 2; Length 452;

Best Local Similarity 36.4%; Pred. No. 6.3e-20;

Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAQAPALSDQDFADFPALPLDPSAMVAQA-----PQVNVNTKLGYNNAV 77

120 GAGPVTGPAASVPAANM-----PSGVEQVAVKVPVSVVMTDLGRQSE-- 164

QY 78 AGTGVIDPNGVLTNNHVIAGA-----TDINAFVSGSGQTVGVVGVYDR 123

155 EGSVILSADGLITNNHVAVAAPGPGGGLSPKTTVTFF---DGRFASFTVVGADP 221

QY 124 TDVAVQLRQAGGLPSAAICGG--VAVGEPVAVMGNSGGGTPPRAVGRVVALGQTVQ 181

222 TSDIAVVRVQISGLTPTITMGSSADLRVGPVAVGSPGLAGT--VTSGIVSALNRVPS 279

QY 182 ASDSLTGAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGMTAA-----SDNFQLSQG 235

280 TTGE-SGNQNTVLDALQTDAAINFGNSGGALVNMGGQLVGNVSAIATLGADSGDAQSGSI 338

QY 236 GQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGN-GARVQRVVGSAAPAA 294

339 GLGFAIPVDQAKRIADELISTG--KATH-----ASLGVQVATDKGTPGAKVMDVWAGGAA 392

QY 295 SLGISTGDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAE 351

393 NAAVFKGVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDQSGSSRTVQVTLGK 449

Db 393 NAAVFKGVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDQSGSSRTVQVTLGK 449

RESULT 7

S74643

proteinase hhoA (EC 3.4.-.-) - Synchococystis sp. (strain PCC 6803)

N: Alternate names: protein sill1679

C: Species: Synchococystis sp.

A: Variety: PCC 6803

C: Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002

C: Accession: S74643

R: Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996

A: Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchococystis sp.

A: Reference number: S74322; MUID:97061201; PMID:8905231

A: Accession: S74643

A: Status: nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-394 <KAN>

A: Cross-references: EMBL:D90900; GB:AB001339; NID:gl651768; PIDN:BAAL6795.1; PID:gl651861

A: Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C: Genetics:

A: Gene: hhoA

C: Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypsin

C: Keywords: hydrolase; proteinase

Query Match 21.2%; Score 382; DB 2; Length 394;

Best Local Similarity 31.6%; Pred. No. 2e-15;

Matches 116; Conservative 50; Mismatches 117; Indels 84; Gaps 14;

QY 13 WLISVLAAVGLGLATAP-----AQAPPA-----LSQDRFADFPALPLDPSAMVA 57

13 YLLAFAVGTAFGIANLPHAVAAADLLPAPVITAQASVPLTSESFV-----AAAVS 63

QY 58 QVAPQVNVNTK-----LGYNNAV-----GAGTGIVIDPNGV 90

64 RSGPAVVRIDTETVTRTRTDLPDLPFFQEPFGRFFPPPRERRTAGGSGFIIDNSGII 123

QY 91 LTNNHVIAGATDINAFVSGSGQTVGVVGVYDRTDQVAVLQRGAG-GLPSAAIG--GGV 147

```

Db 124 LTNHAVVDGASKV-VTLRDGRTDGGVQRTGDEVTDLAVKIEPQSGALPVAPLGTSSNL 182
QY 148 AVGEPVAMGNSGGGGTTPRAVGRVVALG--QTVQASDSLTGAETLNGLIQFDAAIQP 205
Db 183 QVGDWAIAGVNPVGLDNT-----VTLGIISTLGRSAAQAGIPDKRVEFIQTDAAINP 234
QY 206 GDSGPPVNLGQVVGWMTAASDNFQLSQGGFAIPIGQAWATAGAIRSGGSPVTHIG 265
Db 235 GNSGGLNAGGEVLGINTA-----IRADATGIGFAIPIDQAKAIONTLAAGGTVPHPYIG 290
QY 266 PTAFLGLGV-----VDNNGN-----GARVQRVVGSAAPASLSLGTSGDVTAVDGP 311
Db 291 -VQWNNITVDQAQQNRRNPSPFIPEVDGILVMKVLPGTPAERAGIRRGDVIIVADGTP 349
QY 312 INSATAM 318
Db 350 ISDGARL 356

RESULT 8
AG2150
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2150
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001.
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2150
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074457.1; PID:g17131851; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Gene: alk2758
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
Query Match 20.6%; Score 372; DB 2; Length 407;
Best Local Similarity 30.3%; Pred. No. 7.8e-15;
Matches 125; Conservative 59; Mismatches 141; Indels 88; Gaps 15;
QY 7 RSLR--MSWLLSVLAAGVGL---GLATAPAQAAP-----PALSDQRFADFPALPLD 51
Db 10 RSTRLQGLTHLAFIGVVLTVSSLSVLPSQAEPAPNPSTTGSAPELVAQSQSTAIVGN 69
QY 52 PS---AMVAQVAPQVNNIKLGNNAV-----GAGTG 81
Db 70 SSFVTAANRVGSAVRIDTERTITRRVDFLEDPPFRFRFGFGQQLPPEQMRGLGSG 129
QY 82 IVIDPNGVLTNNHVIAGATDINAFSGQTVGVVDVQVDRDQVAVLQLRGAGLPSA 141
Db 130 FIIDKSLIITNAHVVDKADRVTV-RLKGRSFDGKVGQIDEVTDLVAVKINAGNSLPVA 188
QY 142 AIG--GGVAVGEPVAMGNSGGGGTTPRAVGRVVALG--QTVQASDSLTGAETLNGLI 197
Db 189 PLGSSNNVQVGDWAIAGVNLPLGPDNT-----VTLGIVSTLKRSSAQVGITDKELDFI 240
QY 198 QFDAAIQPGSGPVNGLGVVGMATAASDNFQLSQGGFAIPIGQAWATAGAIRSGG 257
Db 241 QTDAAINFGNSGGLNDKGEVIGINTA-----IRADAMGIGFAIPIDKAKATATQERDG 296
QY 258 GSPVTHIGPTAFGLGV-----VDNNG-----NGARVQRVVGSAAPASLSGI 298
Db 297 --KVAH-----PYLGVQMATLTPELAQQNNIDPNSAFAPVENVGLVIRVVPNSPAANAGI 350
QY 299 STGDVITAVDGAIPINATAWADALNHGHPGDVIVSNWOTKSGGTRTGNVTIAE 351
Db 351 RRGDVLQVDGQAITAEQLQNNVENSRLGQALQVRLQ-RGNQTQQLSVRTAE 402

```

```

RESULT 9
S77538
serine proteinase (EC 3.4.21.-) htrA - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein sir1204
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
C:Accession: S77538
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S77538
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-452 <KAN>
A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PIDN:BA017385.1; PID:g165246;
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
C:Keywords: hydrolase; serine proteinase
Query Match 19.7%; Score 355.5; DB 2; Length 452;
Best Local Similarity 29.2%; Pred. No. 8.1e-14;
Matches 116; Conservative 65; Mismatches 129; Indels 87; Gaps 16;
QY 16 SVLAAGVGLGATAPAAQAAPPALSDQRFADFPALPLDPSAMTAQV----APQVNNINTKLG 71
Db 84 SVISPLVTNQSIAPANESLATNLQSRLS-----PREPSNFVVDVVESTGPAVVRAINAQKT 138
QY 72 YNNAV-----GAGTGIVIDPNGVLTNNHVIAGATDIN 104
Db 139 VKSQVPQAFNPFFLQRFEGSQMPMPNERNVQRGTSGFVSNQDKIFTNAHVVDGADDEV 198
QY 105 AFSVSGSGQTVGVVDVQVDRDQVAVLQLRGAGLPSAAIGGG--VAVGEPVAMGNSGGQ 162
Db 199 V-TLKDGRSFGRVNGSDPSTDVAVVKIE-AGDLPFTVALGSDHLQVGWEAIAIGNPLGL 256
QY 163 GGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGSGGPVNVGLGVVGM 222
Db 257 DNT--VTTGILSATGR--RSAD--IGVDPKRVEFIQTDAALNPGNSGGLNADGQVIGM 310
QY 223 NTAASDNFQLSQGGFAIPIGQAWATAGAIRSGGSPVTHIGPTAFGLGVVDNN--- 278
Db 311 NTAIQNAQ-----GIGFAIPINKAQEIAQQLIATG--KVEH-----AYLGIQMTMTTELQ 360
QY 279 -----NGGARVQRVVGSAAPASLSLGTSGDVTAVDGAIPINATAWADALNHG 325
Db 361 SQIRQETGMNTPVDKVVIMQVMPNSPAATAKLEQGDVLSLQSQPVENAEQVQSLVGLK 420
QY 326 HPGDVISVNWOTKSGGTRTG---NVTLAEG-----PP 354
Db 421 AVGDEVEL-----GILRNQQQLNLTVTIGALPSAPP 451

RESULT 10
H97199
htrA-like serine protease (with PDZ domain) [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Dec-2002
C:Accession: H97199
R:Noelling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001.
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:121359325
A:Accession: H97199
A:Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-433 <KUR>
A:Cross-references: GB:AE001437; PIDN:RAK80387.1; PID:G15025449; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 19.6%; Score 354; DB 2; Length 433;
Best Local Similarity 32.0%; Pred. No. 9.4e-14;
Matches 101; Conservative 51; Mismatches 110; Indels 54; Gaps 10;

QY 50 LPSAMVAQVAPVNNITKL-----GYNNAVAGTGIVIDPNGVLT 92
Db 122 LTVSQIVKXVPAVVGSTKTTVTQNDFFSGSSNGSSQTQEGMGSGIIFNNDGYILT 181

QY 93 NHVIAAGATDINAFSGSGTGVVGVYDRTQDVAVLQRLGAGGLPSAAIGGVA--V 149
Db 182 NYHVIKADKI-AVILNNKKEVSAKVYNDENDIAVIXTGSFTVPGVAELGSSASLV 240

QY 150 GEPVWAMNSGGGGTFRAPVGRVVALGQTQVQASDLSLTGABETLNGLIQFDDAIIQPGDSG 209
Db 241 GDSVVAIGNPLGKEFLGTGTGTVVSAVNEVAVSE--GQKQT---YIQTDAAINPGNSG 294

QY 210 GPVWNGLGQVGNWNTAASDNFQSQGQ---GPAIPGQAMAIAGQIRSGGSPVTHGP 266
Db 295 GPLVNSFGQVVGINS-----KISENGVEGIGFSIDITVKSQIKQLSK-----P 339

QY 267 TAPLGLG--VVD-----NNGGARVQVVGSAAPASLGISGTGVDITAVDGAIPNSATA 317
Db 340 ILMLGISGEAVDSTAEQHNIPGVVIEQIQDFSSAQKAGMQGVDTTFDKKVTSTSD 399

QY 318 MADALNGHHPGVISV 333
Db 400 IDSIRKSHNSGDTVQV 415

RESULT 11
T35287
probable secreted proteinase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Dec-2002
C:Accession: T35287
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21574
A:Accession: T35287
A:Status: preliminary; translated from GB/EMBL/DD5J
A:Molecule type: DNA
A:Residues: 1-362 <SEE>
A:Cross-references: EMBL:AL096872; PIDN:CAB51255.1; GSPDB:GN00070; SCOEDB:SC5F7.30
A:Experimental source: strain A3(2)
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 19.1%; Score 345; DB 2; Length 362;
Best Local Similarity 22.6%; Pred. No. 2.6e-13;
Matches 125; Conservative 49; Mismatches 154; Indels 56; Gaps 14;

QY 1 MNSRRSLRMSWLLSVLAAVGLGATA-----PQAAPPALSQDR 41
Db 1 MDTSRTRLRLAPVAVPACV-LLLATGCDAGTDRGSGSAREGDTACAAAPRAASEL 59

QY 42 FADFPALPDPASMAQVAPQVNNITKLGYNNAVAGTGIVIDPNGVLTNNHVIAGAT 101
Db 60 EADY-----ERVIKDPLSVVQIQ-----GDSLGSVVYDDKGVVTVNAHVVG--- 103

QY 102 DINAFSGSGTGVG---VDVGVYDRTQDVAVLQ-RCAGGLPSA--AIGGVAVGPVVA 155
Db 104 DAKSFRVTTARTEGALTAKLVSSYPEDQLAVIKLDKVPCEGMRAARFADSAKVEVGIVLA 163

QY 156 MNSGGGGTFRAPVGRVVALGQTQVQASDLSLTGABETLNGLIQFDDAIIQPGDSGGPVNG 215

Db 164 MGSPGLGLSS--VTQGIVSATGRVTTEGSGGGTGATIANMVQTSAAINPNSGALVNL 221
QY 216 LGQVVGWNTAASDNFQSQG---GQGFAPIGQAMAIAGQIRSGQ---GSPVTHIGPTAF 269
Db 222 DQVIGIPTLLAATDPLGDSNAPGIGFAPASMTVTTVAGQIVLRDKYVTDGSRALGITA- 280

QY 270 LGLGVDDNNGN--GARVQVVGSAAPASLGISTGDTAVDGAIPNSATAMADALNGHHP 327
Db 281 --RTVDDSYRPAAGAAVVEVSDGGAADDAGLRPGDVLVTKLGDITDITITTSSEALASMRP 338

QY 328 GDVISVNWQTKSGGTRTGNVTLAE 351
Db 339 GDRTKVTY-TRDGKHTAEVTLGE 361

RESULT 12
AG0433
proteinase (EC 3.4.21.-) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG0433
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0433
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-457 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92795.1; PID:G15981488; GSPDB:GN00175
A:Gene: degQ
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 19.0%; Score 343; DB 2; Length 457;
Best Local Similarity 31.2%; Pred. No. 4.4e-13;
Matches 122; Conservative 53; Mismatches 126; Indels 90; Gaps 18;

QY 12 SWLLSVLA-AVGLGLATAP---AQAAPPALSQDRFADFPALPDPASMAQVAPQVNNI- 66
Db 5 SLLLSALAI SVGLGLASVPMVSAALPAVAGQ---TLPSL-----APLEXVLPVAVSVH 57

QY 67 -----NTKLGYNNA---VGAGTGIVID-PNGVLTNNHVIAGAT 101
Db 58 VSGSAQQOQLPEEFKFFFGPNAPSGKESRPFEGLOGSVIINAEKGYLTNNHVINAD 117

QY 102 DINAFSGSGTGVVGVYDRTQDVAVLQRLGAGGLPSAAI--GGGVAVGEPVAVMGS 159
Db 118 KIRV-QINDGREYDAKLLGRDEQTDIALQLTDAKNLTAIXADSNLRFVGDFAVAGNP 176

QY 160 GQGGTTPRAVGRVVALGQTQVQASDLSLTGAE-ETLNGLIQFDDAIIQPGDSGGPVVNGLQ 218
Db 177 FGLGQT--ATSGII-SALGR-----SGLNLEGIENFIQTDASINRNSGALVLDGE 226

QY 219 VVGWNTAASDNFQSQGQ---GPAIPGQAMAI-----GOIRSG-----GSPVTH 263
Db 227 LIGINTAI-----IAPGGNIGIIPAISNNAQLNSQOLIIEFGEVKRGLGIRGSEMTAD 281

QY 264 IGPTAFLGLGVDDNNGGARVQVVGSAAPASLGISTGDTAVDGAIPNSATAMADALN 323
Db 282 TAKAFNI-----DAQRGAFVSEVLPKSAAAKAGIKPGDVLISVDGKKISSFAELRAKV 335

QY 324 GHHPGDVLSVNWQTKSGGTRTGNVTLAEPPP 354
Db 336 TTGPGKTIKIG-----LLREGKP 353

RESULT 13
AB2581

periplasmic proteinase XP2241 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: AG2581
R:anonymous. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <SIM>
A:Cross-references: GB:A004037; GB:A003949; NID:g9107394; PIDN:AAF85040.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XP2241
C:Superfamily: Helicobacter serine proteinase

Query Match 19.0%; Score 341.5; DB 2; Length 514;
Best Local Similarity 31.6%; Pred. No. 6.1e-13;
Matches 123; Conservative 55; Mismatches 128; Indels 83; Gaps 17;
QY 3 NSRRSRWSMLSV---LAAGVGLATAPAAQAP--PALSDQRFADFPALPL-----DP 52
Db 2 NSRIHT-RCFGLLAITPLAACGQHNSPTTAPSTPIVTP---STTPAQVLVAGLPDF 57
QY 53 SAMVAQAPQVNT-----KLGYNNAV----- 76
Db 58 TQLVDQVGPVNTETITRKKVGRGILPNDIDPEFFRRFFGPDQFQMPNQRGQDDDE 117
QY 77 -----GAGTGIVDPNGVLTNNHVTAGATDINAFSGSGQTYGVVDYDRDTQDVAVL 130
Db 118 GGIACRGWGSGFIISKDGYLTNNHVTAGSEV-TKLTDRREFKAKIGSDSQYDVALL 176

QY 131 QLRGAGGLPSAAGGGVAV--GEPVAMGNSGGQGTTPRVPGRVVALGQTVQASDSLGT 188
Db 177 KI-DAKNLPTVRIGDSSLSKGGVVAIGSPFGLDHSVTA--GIVSALGRS-----TS 226
QY 189 ABEETLGLIQDAAIQDSDSGPVVNLGQVGMNTAASDNFOLSQ--GQGFAPIGQA 246
Db 227 DQQRVFPFQTDVPEINGNSGGPLNTRGEVGIN---SOIFSASGYMGSFAIPINLA 283
QY 247 MAIAGQIRSGG---GSFTVHIQPTAPL---GLGVVDNNGNGARVQVVGSAASLGIS 299
Db 284 INAAEQIRKTKQVRSMLGVGEIPDALKAQGLGLPDS--RGALVNNIPHSFAAKAGIE 341
QY 300 TGDVITAVDGAPNSATAMADALNGHHPG 328
Db 342 VGDVIRSWGKVISSFSDLPLPGMPPG 370

RESULT 14

AD1894

serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD1894

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1894
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-429 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072659.1; PID:g17130047; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0702
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryp

Query Match 18.6%; Score 335.5; DB 2; Length 429;
Best Local Similarity 29.2%; Pred. No. 1.1e-12;
Matches 110; Conservative 59; Mismatches 137; Indels 71; Gaps 14;
QY 18 LAAVGLGLATAPAAQAPPALSDQRFADFPALPLDPSAMVAQVAPQVYNNITKLGVNNAV- 76
Db 67 LAQKASDLAVSRVDAAPP-LGNNTDPNF-----VTQVQVRGPAVVRIEASRTVTSKLP 119
QY 77 -----GAGTGIVDPNGVLTNNHVTAGATDINAFSGSG 111
Db 120 AEFNDPFFRRFFGSQLPQOQERVQRTGSGFLISADGSLTNAHVVDGADTVRVI-LKDG 178
QY 112 QTYGVVDYDRDTQDVAVLQRLGAGLPSAAGGG--VAVGEPVAMGNSGGQGTTPRAV 169
Db 179 RSFGKVLGTDLNLTDAVYVVIQ-ANNLPTLVGNSDQLQPCQWAIATGNPLGLDNT--VT 235
QY 170 PGRVVALGQTVQASDSLGTGAEETLGLIQDAAIQDSDSGPVVNLGQVGMNTAASDN 229
Db 236 TGIISATGRT-----SNQIGAPDKRVEYIQTDAINPNSGGPFLNRYRGEVGMNTAIQ 291
QY 230 FQLSGGGGFAPIQGAQVAIAGQIRSGGSGPTVHIGTAFGL-----GVV 275
Db 292 AQ-----GLGFAIPKTAQRISNQLIATGKVQHPYLG-IQVGLTPQIKQINSDPNSGLT 346
QY 276 DNNNGARVQVVGSAAPASLIGSTGVDVITAVDGAPINSATAMADALNGHHPGDIVSN- 334
Db 347 VDRDKGVLVVRVLPNSPAARAGLRAGDVIQKLNGQAVTDASNVORAVENAQVGLQLEL 406
QY 335 WTKSGGTRTG-NVTLA 350
Db 407 W-----RNGRNLNLA 416

RESULT 15

AD2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD2057
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073707.1; PID:g17131098; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2008
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryp

Query Match

18.4%; Score 331; DB 2; Length 416;
Best Local Similarity 31.5%; Pred. No. 2e-12;
Matches 92; Conservative 57; Mismatches 115; Indels 28; Gaps 10;

QY	77	GAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVGVDRTOQVAVLQIRGAG	136
Db	133	SGSGGFIISSSGOILTNAHVVDGDEVTV-TLKDGSRFDGKVLGEDPVTDAVAVIQI-NAN	190
QY	137	GLPSAAIGGGVAV--GEPVVMGNSGGQGGTPRAVFGRVVVALGQTVQASDSLTGAETIN	194
Db	191	NLFTVAVGNSEVLQPGGEAVIAIGNPLGLNNS--VTSGIISATGR-----SGSDIGASDXRV	244
QY	195	GLIQFDAAIQGDSGGPVVNGLGQVVGMNTAAASDNFQLSQGGGGAIPIGQAMAIAGQIR	254
Db	245	DYLQTDAAINPGNSGGPLNARGQVIGMNTAIIQGAQ-----GLGPAIPINTVKVQSQELI	300
QY	255	SGGSGPTVHIG-PTAFILGLGVD--NNGNGARVQ-----RVGSAAPAASLIGISTG	301
Db	301	TQGVKVBHPYLGVMATLTQVKERINERFGDRINITADRGVLLVRIVPGSPAANAGLRPG	360
QY	302	DVITAVDGAIPINSATAMADALNGHHPCDVISVNWQTKSGGTRTGNTLAEGP	353
Db	361	DIQSINNGSVTTVEQVKIVENSQIQPLQIQIE--RNGQTTQVNVSPAPLP	411

Search completed: June 22, 2004, 17:24:47
Job time : 10.9325 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 5.9081 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802

Sequence: 1 MSNSRRRSURWSLLSVLAA.....QTKSGGTRTGNVTLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	350.5	19.5	458	1	YVTA_BACSU
2	328	18.2	355	1	DEGS_SCOLI
3	326	18.1	455	1	DEGO_SCOLI
4	322	17.9	475	1	DEGP_SALTY
5	321	17.8	413	1	HTRA_LACHE
6	317.5	17.6	474	1	DEGP_ECOLI
7	310	17.2	408	1	HTRA_LACLA
8	306	17.0	513	1	DEGP_BRUME
9	305	16.9	513	1	DEGP_BRUSU
10	295.5	16.4	449	1	HTRA_BACSU
11	295	16.4	466	1	HTRA_HAEIN
12	291	16.1	504	1	DEGP_RHIME
13	290	16.1	503	1	DEGP_BARHE
14	289.5	16.1	437	1	DEGI_ARATH
15	288	16.0	497	1	DEGP_CHLTR
16	286.5	15.9	497	1	DEGP_CHLMU
17	286	15.9	478	1	DEGP_BUCAI
18	283	15.7	508	1	DEGP_RICCN
19	282	15.6	488	1	DEGP_CHLPN
20	276.5	15.3	400	1	YVTA_BACSU
21	276	15.3	448	1	DEGS_ARATH
22	272.5	15.1	478	1	DEGP_BUCAP
23	263	14.6	458	1	HRA2_HUMAN
24	258	14.3	513	1	DEGP_RICPR
25	254	14.1	458	1	HRA2_MOUSE
26	244.5	13.6	340	1	DEGS_HAEIN
27	238.5	13.2	453	1	HRA3_HUMAN
28	235	13.0	480	1	HRA1_HUMAN
29	231.5	12.8	476	1	HRA4_HUMAN
30	225.5	12.5	460	1	HRA3_MOUSE
31	225	12.5	480	1	HRA1_MOUSE
32	178.5	9.9	321	1	SPPA_ARATH
33	173	9.6	630	1	Y4BT_RHISN

RESULT 1

YVTA_BACSU
ID YVTA_BACSU STANDARD; PRT; 458 AA.

AC Q9R9IL; O35021; O35039;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Probable serine protease yvta (EC 3.4.21.-).

GN YVTA OR VYTB OR BSU33000.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=20158875; PubMed=10692364;

RA Noone D., Howell A., Devine K.M.;

RT Expression of ykdA, encoding a Bacillus subtilis homologue of HtrA,

is heat shock inducible and negatively autoregulated.;

RL J. Bacteriol. 182:1592-1599(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=98015415; PubMed=9353931;

RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;

"Sequencing of regions downstream of addA (98 degrees) and citG (289

degrees) in Bacillus subtilis.;"

RL Microbiology 143:3305-3308(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,

Ghim S.Y., Glasser P., Goffeau A., Golightly E.J., Grandi G.,

Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,

Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

Kurita K., Lapilus A., Lardinois S., Lauber J., Lazarevic V.,

Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

Noore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

Presecan E., Puic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie F.,

Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,

Sorokin A., Tacconis E., Takagi T., Takahashi H., Takemaru K.,

Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,

Tosato V., Uchiyama S., Vanderbol M., Vannier F., Vassarotti A.,


```

RC RX MEDLINE=0157:H7 / RIMD 0509952;
RA HAYASHI T., MAKINO K., OHNISHI M., KUROKAWA K., ISEHII K., YOKOYAMA K.,
RA HAN C.-G., OHSUBO E., NAKAYAMA K., MURATA T., TANAKA M., TOBE T.,
RA IIDA T., TAKAMI H., HONDA T., SASAKAWA C., OGASAWARA N., YASUNAGA T.,
RA KUHARA S., SHIBA T., HATTORI M., SHINAGAWA H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
RX MEDLINE=88105815; PubMed=3322223;
RA VOGEL R.F., ENTIAN K.-D., MECKE D.;
RT "Cloning and sequence of the mdh structural gene of Escherichia coli
RT coding for malate dehydrogenase.";
RL Arch. Microbiol. 149:36-42(1987).
RN [7]
RP IDENTIFICATION
RA BAZAN J.F., FIETTERICK R.J.;
RT "Structural and catalytic models of trypsin-like viral proteases.";
RL Semin. Virol. 1:311-322(1990).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15661; AAC43993.1; -
DR EMBL; U32495; AAC44006.1; -
DR EMBL; U18997; AAA58037.1; -
DR EMBL; AE000402; AAC76267.1; -
DR EMBL; AE005551; AAG58363.1; -
DR EMBL; AF002564; BAB37531.1; -
DR EMBL; M24777; -; NOT_ANNOTATED_CDS.
DR PIR; D91142; D91142.
DR PIR; JC6052; JC6052.
DR MEROPS; S01.275; -.
DR EcoGene; EG11652; hhoB.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 355
FT DOMAIN 281 326 PDZ.
FT ACT_SITE 96 96 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 126 126 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 253 253 R -> A (IN REF. 6).
FT CONFLICT 307 307 V -> E (IN REF. 6).
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;

Query Match
Best Local Similarity 30.7%; Pred. No. 1.1e-11; Length 355;
Matches 110; Conservative 47; Mismatches 135; Indels 66; Gaps 13;

QY 15 LSVLAAVGLG-----LATAPAAAPPALSDRFADFPALPDPSAMVAQVAPQVWIN 67
Dy 3 VKLLRSVAIGLIVGALLVAMPSLRLNPLSPQFDSITDTPASYNIAVRRAPVNV- 61
QY 68 TKLGYNNAVGA-----GTGIVDPNGVLTNNHVIAGATDINAFSGVSGQTYGV 116

```

```

Db 62 -----YNRGLNTNSHQNLEIRTLGSGVIMDQRGVIIITKNKVINDDAQI-IVALQDGRVFEA 116
QY 117 DVVGYDRTPQDVAVLOLRGAGGLPSAIGGGVA--VGEPPVAMGNSGGGGTTPRAVPGRVV 174
Db 117 LLVGSDSLTDLAVLKINATGGLFTIPINARVPHIGDVLAINP-----Y 162
QY 175 ALQGTV-QASDSLUGAETLN-----GLIOFDAIOPGDSGGPVNGLGVGMNTRASD 228
Db 163 NLGQITQTGIISATG-RIGLNPTGRQNFLOTDSINHNSGGALVNSLGLMGTLSFD 221
QY 229 --NPLSQGGGGFAIP-----IGQAMALAGQIRSGGGSPVTHIGTAFGLGVV 275
Db 222 KSDGETPEIGFAIPQLATKMKLIRGVRVIGIGR-----EIAPLHAQGGGI- 276
QY 276 DNNNGGARVQVRVVGSAASLGISTGDTVITAVDGAIPNSATAMADALNHHHPGDSIVS 333
Db 277 -DQLQGIWVNEVSPDGAANAGIQVNDLIISVDNKPALSALETMDQVAEIRPGSVIEP 333

RESULT 3
DEQG_ECOLI
ID DEQG_ECOLI STANDARD; PRT; 455 AA.
AC P39099;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease degQ precursor (EC 3.4.21.-).
GN DEQG OR HHOA OR B3234.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Bass S., Gu Q., Goddard A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
[2]
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
RC STRAIN=K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;
RA Waller P.R., Sauer R.T.;
RT "Characterization of degQ and degS, Escherichia coli genes encoding
RT homologs of the DegP protease.";
RL J. Bacteriol. 178:1146-1153(1996).
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15661; AAC43992.1; -
DR EMBL; U32495; AAC44005.1; -
DR EMBL; U18997; AAA58036.1; -
DR EMBL; AE000402; AAC76266.1; -

```

```

DR PIR; JC6051; JC6051.
DR MEROPS; S01.274; -.
DR SWISS-2DPAGE; P39099; COLI.
DR EcoGene; EGI2612; degQ.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SMO0228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 455
FT DOMAIN 258 349
FT DOMAIN 355 447
FT ACT_SITE 109 139
FT ACT_SITE 139 139
FT ACT_SITE 214 214
FT ACT_SITE 214 214
SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;

Query Match 18.1%; Score 326; DB 1; Length 455;
Best Local Similarity 30.4%; Pred. No. 1.9e-11;
Matches 119; Conservative 61; Mismatches 117; Indels 94; Gaps 19;

QY 14 LLSVLA-AVGLGL-ATAPQAAPPALSSQDRFADFPALPLDPSAMVAQVAVQVNVNT--- 68
D 14 LLSVLA-AVGLGL-ATAPQAAPPALSSQDRFADFPALPLDPSAMVAQVAVQVNVNT--- 68
Db 7 LLSALALSVGLTSLASQFVAVIPGVADQ-AFLPSL-----APMLEKVLPAVVSVRVEGT 61
QY 69 -----KLGYNNAV-----GAGTGVIDPN-GVVLTNHVIAGATDINAFS 107
D 69 -----KLGYNNAV-----GAGTGVIDPN-GVVLTNHVIAGATDINAFS 107
Db 62 ASQGQKIPKFKFGDGLDQPAQPFGLGSGVLIINASKGYVLIINHVIINQAQKI-SIQ 120
QY 108 VSGSGTGVGVVGVVDRVQDVAVLQLRGAGGLPSAAIGGG--VAVGPFVVMGNSGGQGT 165
D 108 VSGSGTGVGVVGVVDRVQDVAVLQLRGAGGLPSAAIGGG--VAVGPFVVMGNSGGQGT 165
Db 121 LNDGREFDAKLISDQSDIALLIQINPSKLTQIATADSKLRVGVFAVAVGNPFLGT 180
QY 166 PRAVGRVVALGTVQASDLSLTAQE-ETLNLGLQFDAAIQPDGSGGVVNGLGQVVGWMT 224
D 166 PRAVGRVVALGTVQASDLSLTAQE-ETLNLGLQFDAAIQPDGSGGVVNGLGQVVGWMT 224
Db 181 --ATSGIVSALGR-----SGNLGLENFQTDASINRNGSGGALLNLGELIGINT 230
QY 225 AASDNFQLSGGQ---GRAPIGQAWATA-----GQIRSGGSGPTVHIGTAFGLGV 274
D 225 AASDNFQLSGGQ---GRAPIGQAWATA-----GQIRSGGSGPTVHIGTAFGLGV 274
Db 231 AI-----LAPGGGSGVIGFAIPSNMARTLAQQLIDFGEIKRG-----LGIK 273
QY 275 VDNNGN-----GARVQRVVGSAAPAASLIGTGVITAVDGAIPNSATAMADALN 323
D 275 VDNNGN-----GARVQRVVGSAAPAASLIGTGVITAVDGAIPNSATAMADALN 323
Db 274 TENSADIAKAFALDVQRFVSVLPQSGSAKAGVKAGDIITSLNGKPLNSFAELRSRIA 333
QY 324 GHHPGDVISVNWQTKSGGTGTG-----NVTLL 349
D 324 GHHPGDVISVNWQTKSGGTGTG-----NVTLL 349
Db 334 TTEPGT-----KVYLGLENGKPLEVEVTL 358

RESULT 4
DEGP_SALTY STANDARD; PRT; 475 AA.
AC P26982;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DSGP OR HTRA OR PTD OR STM209.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C5;

```

```

RX MEDLINE=91251770; PubMed=1645840;
RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
RA Ali T., Miller I., Hormaeche C.;
RT "The role of a stress-response protein in Salmonella typhimurium
RT virulence.";
RL Mol. Microbiol. 5:401-407(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Bante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
CC SPECIFICITY WITH HHOA/DEGO.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X54548; CAA38420.1; -.
CC EMBL; AE008704; AAL19173.1; -.
CC PIR; S15337; S15337.
CC MEROPS; S01-273; -.
CC StyGene; SG10173; degP.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SMO0228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 475
FT DOMAIN 281 372
FT DOMAIN 378 467
FT ACT_SITE 132 132
FT ACT_SITE 162 162
FT ACT_SITE 237 237
FT DISULFID 84 96
FT DISULFID 84 96
SQ SEQUENCE 475 AA; 49315 MW; 86E685BF3C1A289F CRC64;

Query Match 17.9%; Score 322; DB 1; Length 475;
Best Local Similarity 29.1%; Pred. No. 3.2e-11;
Matches 118; Conservative 55; Mismatches 126; Indels 106; Gaps 16;

QY 20 AVGLGLATAP--AQAAPPALSSQDRFADFPALPLDPSAMVAQVAVQVNVNT----- 68
D 20 AVGLGLATAP--AQAAPPALSSQDRFADFPALPLDPSAMVAQVAVQVNVNT----- 68
Db 12 ALSGLALSLSLAATSSAMTAQQPSL-----APMLEKWPVSVINVEGTTVNT 67
QY 69 -----KLGYNNAVGA-----GTGVID-NGVV 90
D 69 -----KLGYNNAVGA-----GTGVID-NGVV 90
Db 68 RMPRFQFFQDDSPFQDGSFPQNSPFQCGGNGGNGGQKFMALGSGVITDAKGYV 127
QY 91 LTNHVIAGATDINAFVSGSGTQTVGVVGVYDRVQDVAVLQLRGAGGLPSAAIGGVA-- 148

```

Db 128 VTNHVDNASKV-KV-QLSDGRKFDKAVGVKDPKSDIALIQNPKNLTAKLADSDLR 186
Qy 149 VGEPPVAMNSGGQGTTPAUGRVRVVALGQTVQASDSLGTGAB-ETLNGLIQFDALQPD 207
Db 187 VGDYVAINGNPGLGET--VTSIGVSLGR-----SGLNVENYENFIOTDAINRGN 236
Qy 208 SGGPPVNGLGQVVGWNTA--ASDNFQLSGGGQGFAP-----IQANMAIAGQIRSGGGS 259
Db 237 SGGALVNLNGELIGINTAILAFDGGNI--GIGFAPSNMVKNLTSQMVYGVQVGRG-- 290
Qy 260 PTVHIGTFAFLGLGVVDNNGN-----CARVQRVVGSPAPASLIGSTGDVITA 306
Db 291 -----ELGIMGTELNSELAKAMKVDAQRCAFVSQVNPNSAAKAGIKAGDVITS 339
Qy 307 VDGAPINSATAMADALNGHHPDVISVNWQTKSGGTRTGNVTLAE 351
Db 340 LNKGPISFPAALRAQVGTMPVSKISLG-LLREGKAITVNLLEQQ 383

RESULT 5
HTRA_LACHE
ID HTRA_LACHE STANDARD; PRT; 413 AA.
AC Q9Z4H7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htrA (EC 3.4.21.-).
GN HTRA.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus
OX NCBI_TaxID=1587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=53/7;
RX MEDLINE=99047559; PubMed=9829922;
RA Smets A., Varmann P.K., Palva A.M.;
RT Molecular characterization of a stress-inducible gene from
Lactobacillus helveticus.;
RL J. Bacteriol. 180:6143-6153(1998).
CC -|- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -|- SIMILARITY: Belongs to peptidase family S2C.
CC -|- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ005672; CAA06668.1; -
DR MEROPS; S01.273; -
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00899; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Transmembrane.
FT TRANSMEM 20 40
FT DOMAIN 141 302 CATALYTIC.
FT DOMAIN 305 401 PDZ.
FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 413 AA; 42647 MW; B16B677991C88707 CRC64;

Query Match 17.8%; Score 321; DB 1; Length 413;
Best Local Similarity 31.3%; Pred. No. 3.2e-11;
Matches 89; Conservative 50; Mismatches 121; Indels 24; Gaps 8;
Qy 67 NTKLGVNNAVGAGTGIIV---DPNGVVLNNHVIAGATDINAFSVGSGQTVGVVDVVGYDR 123
Db 119 SSRNGKLETYSSEGVVYMKSNKGIVITNNHVISGDAVQVL-LANGKTVNAKVVGKDS 177
Qy 124 TDQVAVLQLRGAGGLPSAAIIGG--VAVGPPVAMNSGGQGTTPRAVPRVVALGQTVQ 181
Db 178 TTDLAVLSDAKVYVTAQFGDQSKHLEAGQTVIAGVSPLGSEYASTVQGIISAPARTIS 237
Qy 182 ASDSLTGAETLNGLIQFDALQPDGSGPPVNGLGQVVGWNT--AASDNFQLSGGGQGF 239
Db 238 TS---SGNQGT---VQTDAINPNSGGALVNSAGQVIGINSMKLAQSSDGTSGVEGNAP 291
Qy 240 AIPGQAMAIAGQIRSGG--GSPTVHIGTFAFLG-----LGVDNNGNGARVQRVVG 289
Db 292 AIPSNVVTIVNELVKKKITRQLGVVRVIALQIGPEGRSLKIKSNLKNIGVIAFVSR 351
Qy 290 SAPAASLGISTGVDITAVDGAIPNSATAMADALNGHHPGDVISV 333
Db 352 NGSANAGIKSGDVITKVDGKVEDVASLHSLYSHKVGDTNVV 395
RESULT 6
DEGP_ECOLI
ID DEGP_ECOLI STANDARD; PRT; 474 AA.
AC P09376; P15724;
DT 01-MAR-1989 (Rel. 10, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-)
GN DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htrA gene of Escherichia
coli: a sigma 32-independent mechanism of heat-inducible
transcription."; Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
the 2.4-4.1 min (110,917-193,643 bp) region."; Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."; Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen B., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Hayek G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamouisis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RN RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RN RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90323597; PubMed=2165018;
RA Quirk S., Bhattacharjee S.K., Bessman M.J.;
RT "Primary structure of the deoxyguanosine triphosphate
RT triphosphohydrolase-encoding gene (dgt) of Escherichia coli.";
RL Gene 89:13-18(1990).
RN [8]
RN RP SEQUENCE OF 1-16 FROM N.A.
RC MEDLINE=90207273; PubMed=2157212;
RA Wurgler S.M., Richardson C.C.;
RT "Structure and regulation of the gene for dGTP triphosphohydrolase
RT from Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
RN [9]
RN RP SEQUENCE OF 27-39, AND CHARACTERIZATION.
RX MEDLINE=90202693; PubMed=2180503;
RA Lipinska B., Zylcz M., Georgopoulos C.;
RT "The HtrA (DegP) protein, essential for Escherichia coli survival at
RT high temperatures, is an endopeptidase.";
RL J. Bacteriol. 172:1791-1797(1990).
RN [10]
RN RP SEQUENCE OF 27-30.
RC STRAIN=K12 / W3110;
RX MEDLINE=98263247; PubMed=9600841;
RA Wilkins M.R., Gasteliger E., Tonella L., Ou K., Tyler M.,
RA Sanchez J.-C., Gooley A.A., Walsh B.J., Bairoch A., Appel R.D.,
RA Williams K.L., Hochstrasser D.F.;
RT "Protein identification with N and C-terminal sequence tags in
RT proteome projects.";
RL J. Mol. Biol. 278:599-608(1998).
RN [11]
RN RP IDENTITY OF HTRA AND PROTEASE DO.
RX MEDLINE=91222240; PubMed=2052528;
RA Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
RA Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
RT "Protease Do is essential for survival of Escherichia coli at high
RT temperatures: its identity with the htra gene product.";
RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
RN [12]
RN RP DISULFIDE BOND.
RX MEDLINE=22760311; PubMed=12878036;
RA Skorko-Glonek J., Zurawa D., Tanfani F., Scire A., Wawrzynow A.,
RA Narkiewicz J., Bertoli E., Lipinska B.;
RT "The N-terminal region of Htra heat shock protease from Escherichia
RT coli is essential for stabilization of Htra primary structure and
RT maintaining of its oligomeric structure.";
RL Biochim. Biophys. Acta 1649:171-182(2003).
RN CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.

```

```

CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE
CC ICA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGP.
CC -1- SUBUNIT: Multimeric.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: By heat shock.
CC -1- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
CC -1- SIMILARITY: Belongs to peptidase family S2c.
CC -1- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M36536; AAA23394.1; -
CC EMBL; X12457; CAA30997.1; -
CC EMBL; D26562; BAB96738.1; -
CC EMBL; AE000125; AAC73272.1; -
CC EMBL; U70214; AAB08591.1; -
CC EMBL; AE005192; AAG54465.1; -
CC EMBL; AF002550; BAB33588.1; -
CC EMBL; M29555; AAA23717.1; -
CC EMBL; M31772; AAA23660.1; -
CC PIR; E85500; E85500.
CC PIR; E90649; E90649.
CC PIR; S45229; S45229.
CC PDB; 1KY9; 03-APR-02.
CC MEROPS; S01-273; -
CC
CC SWISS-2DPAGE; P09376; COLI.
CC Ecogene; EGI0463; degp.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR01478; PDZ.
CC InterPro; IPR01254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ_2.
CC Pfam; PF00083; trypsin_1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 2.
CC HydroLase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
CC Complete proteome; 3D-structure.
CC SIGNAL
CC CHAIN 1 26
CC CHAIN 27 474
CC DOMAIN 280 371
CC DOMAIN 377 466
CC ACT_SITE 131 131
CC ACT_SITE 161 161
CC ACT_SITE 236 236
CC DISULFID 83 95
CC CONFLICT 10 10
CC CONFLICT 46 46
CC CONFLICT 192 192
CC CONFLICT 467 474
CC SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;

```

```

Query Match 17.6%; Score 317.5; DB 1; Length 474;
Best Local Similarity 29.8%; Pred. No. 5.7e-11;
Matches 112; Conservative 51; Mismatches 108; Indels 105; Gaps 16;

QY 15 LSVLAAGVLGLATAPAAQAAPPALSQDRFA-DFFALPLDPSAMVAQVAPQVNVNTK---- 69
DB 8 LSAL-ALSLGLALSPLSATAETSSATTAAQMPSL-----APMLEKVPFSVVSINVESTT 62
QY 70 -----LGYNA-----VCAGTGVDPN 87
DB 63 VNTPRMRFNQFFGDDSPFCQSGSPFQSSPFCCGGGGGNGGQKQKFMALGSGVLIAD 122
QY 88 -GVVLTNNHVIAGATDINAFSGVGQTYGVVDVGYDRTQDVAVLQRLGAGGLPS--AAIG 144

```

```
Db 123 KGVVTVNNHVVNDVATVVKV-QISDGRKFDKAVKWDKPRSDIALIQIPKNTAKMADS 181
QY 145 GGVAVGEPVVMGNSGGGCTPRAVEGRVVALGQTVQASDSUTGAB-EFTNLGLIFDAAI 203
Db 182 DALRVGDYTVAGNPFGLGET-VTSGIVSALGR-----SGLNAENYENFIQTDAAI 231
QY 204 QPGDSGGPVVNGLGQVVGWNTA--ASDNFQLSGGGGGAIP------IQQAMAIAQIORS 255
Db 232 NRNSGGALVNLNGELIGINTAILAPDGGNI---GIGFAIPNNVKNLTSQWVEYGVQVQR 288
QY 256 GGGSPVTHICPTAFLGLGVVDNNGN-----GARVQRVVGSSAPAAASLGISTDG 302
Db 289 G-----ELGIMGTLSNELAKAMKVDAQRGAFFVSQVLPNSSAAKAGIKAGD 334
QY 303 VITAVDGAIPNSATAM 318
Db 335 VITSLNGKPISSFAAL 350

RESULT 7
HTRA_LACLA
ID HTRA_LACLA STANDARD; PRT; 408 AA.
AC Q9LA06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-) (Htra11).
GN HTRA OR IL2136.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=IL1403;
RX MEDLINE=20177820; PubMed=10712686;
RA Poquet I., Saint V., Seznec E., Simoes N., Bolotin A., Gruss A.;
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis sp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: DEGRADATES ABNORMAL EXPORTED PROTEINS. NEEDED FOR THE PRO-
CC PEPTIDE PROCESSING OF A NATURAL PRO-PROTEIN AND FOR MATURATION OF
CC A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED
CC PROTEINS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (probable).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENMBL; AF155705; AAF61294.1; .
CC EMBL; AE006442; AAK06234.1; .
CC EMBL; H86891; H86891.
CC MEROPS; S01.273.
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
```

```
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO00834; PROTEASES2C.
DR SMART; SMO0228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Transmembrane; Complete proteome.
FT TRANSMEM 6
FT DOMAIN 88 284
FT DOMAIN 302 383
FT DOMAIN 127 127
FT ACT_SITE 157 157
FT ACT_SITE 239 239
FT ACT_SITE 239 239
SQ SEQUENCE 408 AA; 41648 MW; 581B90B5A7DF851 CRC64;

Query Match 17.2%; Score 310; DB 1; Length 408;
Best Local Similarity 31.3%; Pred. No. 1.3e-10;
Matches 93; Conservative 51; Mismatches 119; Indels 34; Gaps 11;

QY 79 GTGIVIDPENG---VVLNNHVIAGTADINAFSGSGQTVGVVDVGYDRTQDAVLAQLRG- 134
Db 108 GSGVIYKSGGDAYVVTNHYHVIAGNSLDVLLSG-GQVKASVVGVDYDTDLAVLKISSE 166
QY 135 -AGGLPSAIGGVAVGEPVVMGNSGGGCTPRAVEGRVVALGQTVQASDSLGAETL 193
Db 167 HVKDVATFADSSKLTIGEPALVAGSPGLSQPANTATATGILSATSRQVTLQE-NGQITNI 225
QY 194 NGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTA---ASDNFQLSGGGGGAIPIQQAMAIA 250
Db 226 NA-IQTDAAINPGNSGGALINIEQVIGITQSKITTEDGSTVEGLGFAIPSDNDVNNII 284
QY 251 QGIRSGGSPVHIGPTAFLGLGVVD-----NNGN-----GARVQRVVGSSAPAA 294
Db 285 NKLEADG-----KISRPA-LGIRMVDLSQLSTNDSQKLFPSSVTGGVGVVSVSGSLPAA 338
QY 295 SLGISTGDIVTAVDGAIPNSATAMADALNGHPGDISVNMWQTKSGGTRTGNVTIAE 351
Db 339 SAGLKAGDVITKVGDTAVTSTQSLYSHNINDTVKVTYY-RDGKNTADVLSK 394

RESULT 8
DEGP_BRUME
ID DEGP_BRUME STANDARD; PRT; 513 AA.
AC Q8VG32;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR EMBL1330.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=1175688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N.H., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyprides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; AE009571; AAL52511.1; -;
 DR PIR; AD3418; AD3418.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 DR Hydroxylase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 513 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 125 299 CATALYTIC.
 FT DOMAIN 300 391 PDZ 1.
 FT DOMAIN 414 500 PDZ 2.
 FT ACT_SITE 152 152 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 182 182 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 513 AA; 53514 MW; C472FEF9DFC6268 CRC64;
 Query Match 17.0%; Score 306; DB 1; Length 513;
 Best Local Similarity 28.2%; Pred. No. 2.7e-10;
 Matches 126; Conservative 43; Mismatches 142; Indels 136; Gaps 18;
 QY 1 MSNRRRLRWLLSVLAAGVGLATAPA-----QAAPALSQDRF 42
 Db 1 MSRRRLRWLLSVLAAGVGLATAPA-----QAAPALSQDRF 42
 QY 43 ADFFPALPLDPSAMVAQVAPQVNNITKL----- 70
 Db 54 AD-----LVEKVRPAVSVRVKQVQETSNRGPQFFGPPGDPQLPDGHPKRRFR 103
 QY 71 -----GYNVAVGATGIVIDPNGVLTNNHVIAGATDINAF 106
 Db 104 DFGMEPRGDSNRNRKANKPRPHERPVAQSGFVISEDGVVVTNNHV-----SDGDAY 160
 QY 107 SV--GSGQTGVGVDRYQDVAVLQRCAGGLPS-----AAIG--GVVAVGEPVWAG 157
 Db 161 TVLDDTDELDAKLIGADPTDLAVLKINA----PRKFFVYVAFGDDNKVRGDNVWVAVG 216
 QY 158 NSGQGGTTPRVPGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQGDSGGPVVNGLG 217
 Db 217 NPFGLGGT--VTSGIVSARGRDIGAG-----PYDDFIQIDAAVKNKSGSPAFDLGG 266
 QY 218 QVVGMTAAADNQLSGGGG--FAPIGQMAIAQIIRSGG-----GSTVHLGP----- 266
 Db 267 EVIGINTAI---FSPSGGSGVIGAFAPSSSTAKGVVDOLIKKGSVERGWTVGVIQPVTKDI 323
 QY 267 TAFILGLGVWNGNGARVQVWGSAPASLIGITGVITAVDVGAPINSATAMADALNGHH 326
 Db 324 AASLGLA---EEKGAIVASPDQDGAAPAKAGIKAGDVITAVNGETQDPRDLARKVANTA 379
 QY 327 PGDIVSVNWQTKSGTGTGNVTIABGP 353
 Db 380 PGEKAALTVMRKNKABEI-NVTITAMP 405
 RESULT 9
 DEGP_BRUSU
 ID_DEGP_BRUSU STANDARD; PRT; 513 AA.
 AC Q44597;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).
 GN DSGP OR HTRA OR BR0611.
 OS Brucella suis, and

OS Brucella abortus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OC NCBI_TaxID=29461, 235;
 RN (1)
 RN SEQUENCE FROM N.A.
 RP SPECIES=B.abortus; STRAIN=2308;
 RX MEDLINE=95165990; PubMed=7861951;
 RA Tatum F.M., Chevillat N.F., Morfitt D.;
 RT "Cloning, characterization and construction of htrA and htrA-like
 mutants of Brucella abortus and their survival in BALB/c mice."
 RL Microb. Pathog. 17:23-36(1994).
 RP (2)
 RP SEQUENCE FROM N.A.
 RP SPECIES=B.suis; STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Seanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.B., Halling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; U07352; AAN70164.1; -;
 DR EMBL; AE014368; AAN29540.1; -;
 DR PIR; I40060; I40060.
 DR TIGR; BR0611; -;
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ
 DR InterPro; IPR001254; Peptidase_S1
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 DR Hydroxylase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 513 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 125 299 CATALYTIC.
 FT DOMAIN 300 391 PDZ 1.
 FT DOMAIN 414 500 PDZ 2.
 FT ACT_SITE 152 152 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 182 182 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 513 AA; 53483 MW; DE1CEF1959472806 CRC64;
 Query Match 16.9%; Score 305; DB 1; Length 513;
 Best Local Similarity 28.2%; Pred. No. 3e-10;
 Matches 126; Conservative 43; Mismatches 142; Indels 136; Gaps 18;
 QY 1 MSNRRRLRWLLSVLAAGVGLATAPA-----QAAPALSQDRF 42
 Db 1 MSRRRLRWLLSVLAAGVGLATAPA-----QAAPALSQDRF 42
 QY 43 ADFFPALPLDPSAMVAQVAPQVNNITKL----- 70
 Db 54 AD-----LVEKVRPAVSVRVKQVQETSNRGPQFFGPPGDPQLPDGHPKRRFR 103

QY 71 -----GYNVAVGAGTGIVIDPNCVLTNNHVIAGATDINAF 106
DB 104 DFCMEPRGDSRSDNRGKANKPRPGRHPRVAGSGFVISEDGYVYNNHV---SDGDAY 160
QY 107 SV--GSGQTYGVVDVYDRTQDVAVLQLRGAGCLPS-----AAIG--GGVAVGEFVAVMG 157
DB 161 TVVLDDGTGLDAKLIGADRTDLAVLKINA-----PKRFEVYVAFGDDNKVRVGDVWVAVG 216
QY 158 NSGGGGTTPRVPGRVVALGQTVQASDSLUTGABETLNGLIQDAAIQGDSGGPVVNGLG 217
DB 217 NPFGLGGT--VTSGIVSARGRDIGAG-----PYDDFIQIDAAVWNGNSGGPFAFDLSG 266
QY 218 QVGVNMTASDNFQLSGQGG--FAIPICQAMAIAGQIRSGG-----GSPVTHVIG----- 266
DB 267 EVIGINTAI--FSPSGGVGTAFAIPSTAKQVVDQLIKGVSERGMVIGVQIQVTKDI 323
QY 267 TAFGLGVVDNNGNGARVQVWGSAPASGLISTGVTITAVDGPINSATAMADALNGHH 326
DB 324 AASLGLA---BEKGAIVASPDQDGAAPKAGI-KAGDVITAVNGETVQDPRDLARKVANIA 379
QY 327 PGDIVSNWQTKSGGRTGNVTLAEGP 353
DB 380 PGKAAALTVWRKNKABEI-NVTIAAMP 405

RESULT 10
HTRA_BACSU
ID HTRA_BACSU STANDARD; PRT; 449 AA.
AC O34358;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease do-like htra (EC 3.4.21.-).
GN HTRA_OR BSU12900.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RT "Sequence of the Bacillus subtilis genome between xlyA and ykoR";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=38044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolojin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serfor P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarctti A.,

Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
"The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis"; 249-256 (1997).
[3]
RN Nature 390:249-256 (1997).
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykxA, encoding a Bacillus subtilis homologue of Htra,
is heat shock inducible and negatively autoregulated";
RL J. Bacteriol. 182:1592-1599 (2000).
[4]
RN TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=11133960;
RA Noone D., Howell A., Colley R., Devine K.M.;
RT "YkxA and yvta, Htra-like serine proteases in Bacillus subtilis,
engage in negative autoregulation and reciprocal cross-regulation of
ykxA and yvta gene expression";
RL J. Bacteriol. 183:654-663 (2001).
[5]
RN TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=21439741; PubMed=11555295;
RA Hyvylaeinen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
RA Vitikainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M.,
RA Kontinen V.P.;
RT "A novel two-component regulatory system in Bacillus subtilis for the
survival of severe secretion stress";
RL Mol. Microbiol. 41:1159-1172 (2001).
CC -!- FUNCTION: May be involved in processing, maturation, or secretion
of extracellular enzymes.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- INDUCTION: Transcription is cssS dependent. Induced by heat shock
during exponential growth and by heterologous amyloses at the
transition phase of the growth cycle. Negatively regulates its own
expression during exponential growth and during heat shock.
CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
of yvta, especially during stress conditions.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AJ002571; CAA05570.1; -;
DB EMBL; Z99110; CAB13147.1; -;
DB PIR; A69643; A69643.
DR MEROPS; S01.273; -;
DR Subtilist; BG12608; htra.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolyase; Protease; Serine protease; Heat shock; Transmembrane;
Complete proteome.
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 67 POTENTIAL.
FT DOMAIN 68 49 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 103 108 POLY-SER.

FT DOMAIN 146 152 POLY-SER.
 FT DOMAIN 348 437 PDZ
 FT ACT SITE 179 179 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 209 209 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 230 290 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 449 AA; 47713 NW; E12B07A9018BE414 CRC64;

Query Match 16.4%; Score 295.5; DB 1; Length 449;
 Best Local Similarity 29.3%; Pred No. 8.9e-10;
 Matches 93; Conservative 53; Mismatches 126; Indels 45; Gaps 11;

QY 53 SAMVAQAPQVNNINKLGYNNA-----VGAGTGIVV-DPNG--VVLITNN 94
 DB 119 SDWEDLSFAIVGITNLQAQNSLSFGSSSDSDSDTESGSGVIFPKKENGKAYIITNN 178
 QY 95 HVIAGATDINAFSVGGQTYGVVDVDRDQVAVLQLRGAGGLPSAIG--GGVAVGEP 152
 DB 179 HVVEGASLKV-SLYDGTETVAKLVGSDSLTDLAVLQISDDHVTKVANFGDSDSLRTGET 237
 QY 153 VVAMNGSGGGGTTPRAVGRYVALGQTYQASDSLTGAETLGLIQFDAAIQPGDSGGPV 212
 DB 238 VIAIGDPLGKLSRTVTCQIVSGVDRTVMS---TSAGETSINVIQDAINFGNSGGLP 294
 QY 213 VNGIQQVVGWMT--AASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGTAF 270
 DB 295 LNTDGKIVGINSKISDDVE---GIGFAIPSNVDKPIABELLSKQIERPVIG-VSML 349
 QY 271 GLGVVDNN-----GNGARVQVVGSAPAAASLGISTGDTVITAVDGAIPINSATA 317
 DB 350 DLEQVPQYQEGTGLFGSQLNKVYIREVASSGPAEKAGLKAEIDILGKKEITGSE 409
 QY 318 MADAL-NGHHPGDVISV 333
 DB 410 LRNLYKDAKIGDTVEV 426

RESULT 11
 ID HTOA_HAEIN STANDARD; PRT; 466 AA.
 AC P45129;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable periplasmic serine protease dc/hnoA-like precursor
 DE (EC 3.4.21.-).
 GN H1259.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=9350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerkavag A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
 RL Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA (PROTEASE DO) AND HHOA.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; U32805; AAC22906.1; -;
 DR FIR; A64113; A64113.
 DR MEROPS; S01.274; -;
 DR TIGR; H1259; -;
 DR InterPro; IPR009003; Cys Ser_trypsin.
 DR InterPro; IPR001478; PDZ_1.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; SM00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 DR Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 KW SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 466 PROBABLE PERIPLASMIC SERINE PROTEASE
 FT DOMAIN 270 361 DO/HHOA-LIKE.
 FT ACT SITE 120 120 PDZ 1.
 FT ACT SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 466 AA; 49434 NW; ED050A00047B5851 CRC64;

Query Match 16.4%; Score 295; DB 1; Length 466;
 Best Local Similarity 28.1%; Pred No. 9.9e-10;
 Matches 108; Conservative 54; Mismatches 124; Indels 98; Gaps 14;

QY 17 VLAAVGLGLATAP-----AQAAPPALSDQRFADFFALPLDPSAMVAQVAVVNNITK 69
 DB 7 VLNSIALGLSVLSTSFVAHVACATLPSFVSSQNSLAP-----MLEKQPAVVVTLSE 58
 QY 70 -----LGNNAVAGAGTGIVDPN-GVVLITN 93
 DB 59 GKAKVDSRSPFLDDIPEEFKFFGDRFAEQPGGSGSKRNFRLGSGVVIINASKGYVLTN 118
 QY 94 NHVIAGATDINAFSVGGQTYGVVDVDRDQVAVLQLRGAGGLPSA--AIGGVAVGE 151
 DB 119 NHVIDGADKIV-QIQDGRPEKAKLVGKDEQSDIALVQLEKPSNLTIKFADSDKLAVGD 177
 QY 152 PVVAMNGSGGGGTTPRAVGRVVALGQTYQASDSLTGAEB-TLNGLIQFDAAIQPGDSGG 210
 DB 178 FTVAGNPFGLGQT--VTSGIVSALGRS-----TGSDSGTENYIQTDAAVNRNGSG 228
 QY 211 PVVAMGLGVGMNTAASDNFOLSGGGG--FAIPIGQAVIA-----GOIRSGGSPV 262
 DB 229 ALVNLGELIGINTAI---ISPSGNGAGIAFAIPSNQASNLVQVILEFGQVRG----- 279
 QY 263 HIGTAFELIGVVDNNGN-----GARVORVVGSAAPASLGISTGDTVAVDGAP 311
 DB 280 -----LLGKGGELNADLAKAFNVSAQQGAFVSEVLPSKSAEAKGLKAGDIITAMNGQK 333
 QY 312 INSATAMADALNGHHPGDVISVNW 335
 DB 334 ISSPAEIRAKIATGAGKEISLTY 357

RESULT 12
 ID DEGP_RHIME STANDARD; PRT; 504 AA.
 AC Q52854;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).

DEGP1 OR DEGP OR R01021 OR SMC02365.
Rhizobium meliloti (Sinorhizobium meliloti).
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
NCBI_taxid=382;
RN [1]
SEQUENCE FROM N.A.
STRAIN=1021;
MEDLINE=96146524; PubMed=8550509;
RA Glazebrook J., Ichige A., Walker G.C.;
RT "Genetic analysis of Rhizobium meliloti bacA-phoA fusion results in
RT identification of degP: two loci required for symbiosis are closely
RT linked to degP";
RL J. Bacteriol. 178:745-752(1996).
RN [2]
SEQUENCE FROM N.A.
STRAIN=1021;
MEDLINE=21396507; PubMed=11481430;
RX Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godtke T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandebol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021".
RN Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).

CC EMBL; U31512; AAC43669.1; ALT_INIT.
DR EMBL; AL591785; CAC45593.1; --
DR MEROPS; S01.273; --
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptide_SIC.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
Complete protosome.
FT SIGNAL 1 26
FT CHAIN 27 504
FT DOMAIN 113 286
FT DOMAIN 287 378
FT DOMAIN 401 491
FT ACT_SITE 140 140
FT ACT_SITE 170 170
FT ACT_SITE 244 244
FT CONFLICT 14 15
FT CONFLICT 39 147
FT CONFLICT 464 504
FT CONFLICT
FT SEQUENCE 504 AA; 53035 MW; D7E82BB9981EA23C CRC64;
Query Match 16.1%; Score 291; DB 1; Length 504;
Best local similarity 33.3%; Pred No 1.8e-09;
Query Match 16.1%; Score 291; DB 1; Length 504;
Best local similarity 33.3%; Pred No 1.8e-09;

Db 220 GVADLLVGKVFAGNPFGLDHT--LTTGVISGLRREI--SSAATG--RPTQDVIOQTA 273
QY 202 AIOFGDSGGVWAGLGWGMNTAASDNFOLSGGQGFAPICQAMAIAGQIRSGGSPT 261
Db 274 AINPGNSGGLLDSGFLICGINTAIYSPGAS--SGVFSIPVDTGIVDQL----- 324
QY 262 VHGPTAFGLGV-----VDNNG-NGARVQVVSAPAAISLIGST-----GD 302
Db 325 VRFKGVTRPILGFKFAPDQVEQLGVSGLLLDAPPSPGAGKAGLQSTKRDGVGRILGD 384
QY 303 VITAVDGAIPINSATAMADALNHHHPGDVSV 333
Db 385 IITSVNGTKVNSGSDLYRILDCKQGVDEVTV 415

RESULT 15
DEGP CHLTR STANDARD; PRT; 497 AA.
AC P18584; O84830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa
DE immunogenic protein) (SK59).
GN DEGP OR HTRA OR CT823.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
CX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar L2;
RX MEDLINE=90337348; PubMed=2379836;
RA Kahane S., Weinstein Y., Sarov I.;
RT "Cloning, characterization and sequence of a novel 59-kDa protein of
RT Chlamydia trachomatis";
RL Gene 90:61-67(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis";
RL Science 282:754-759(1998).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/HR domains.
CC -!- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
CC PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001355; AAC68420.1; -
CC EMBL; M31119; AAA23116.1; -
CC PIR; H71465; H71465.
CC PIR; H71465; P18584; -
CC PHCI-2DPAGE; P18584; -
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ_
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR008256; Peptidase_S1B_Va.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.

PRINTS; PR00839; V8PROTEASE.
SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Repeat; Signal; Antigen;
KW Complete proteome.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 497 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 128 289 CATALYTIC.
FT DOMAIN 290 381 PDZ 1.
FT DOMAIN 394 485 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 497 AA; 53244 MW; 86A5E31BB8438BA CRC64;

Query Match 16.0%; Score 288; DB 1; Length 497;
Best Local Similarity 32.3%; Pred. No. 2.6e-09;
Matches 95; Conservative 46; Mismatches 109; Indels 44; Gaps 12;
QY 79 GTGIVDPNGVLTNNHVIAGATDINAFSVSGSGQTYGVVDVYDRTQDVAVLQLRAGGL 138
Db 127 GTGFIIVSEGGVVTNNHVVEDAGKIHV--TLHDGQYTKIVGLDPTDLAVIKIQ-AEKL 184
QY 139 P5AAIIGGG--VAVGEFVVAMGNSGGQGTTPRAVGRVVALGQT----VOASDSLTAET 192
Db 185 PELTFGNSDQLQIGDWAIAGNPFGLQAT--VTGVISAKGRNQLHIVDFED----- 234
QY 193 LNLGIQFDAATQPCDSGSPVNGLVGMNTASDNFQLSQG--GQGFAPICQAMAIA 250
Db 235 ---FIQTDAAINPGNSGGPLNNGQVIGVNTAIVSG---SGGYIGIGFAPISLMKXVI 288
QY 251 GQIRSGGSPVTHIGTAFGLGVVDN-----NNGARVQVVSAPAAISLIGSTG 301
Db 289 DQLISDGQVTRGFLGT---LQPIDSELATCYKLEKYVYGVALVTDVVKGSFAEKAGLQ 344
QY 302 DVITAVDGAIPINSATAMADALNHHHPGD--VISVNWQTKSGGTRTGNTLAEGP 353
Db 345 DVIVAYNGKEVESLSALRNALSLMMPGTRVVLKI---VREGKTIPIPTVTQIP 395

Search completed: June 22, 2004, 17:16:46
Job time : 7.99901 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 29.0268 Seconds
(without alignments)

3858.816 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802

Sequence: 1 MNSRRRSRLRWSLLVLA.....QTKSGGTRGNVTLAEGPPA 355

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriapi.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.4	355	16 Q07175	Q07175 mycobacteri
2	1792	99.4	355	16 Q70289	Q70289 mycobacteri
3	1306	72.5	361	2 Q50320	Q50320 mycobacteri
4	1272.5	70.6	354	16 Q9CCY9	Q9CCY9 mycobacteri
5	469.5	26.1	446	16 Q8VKA4	Q8VKA4 mycobacteri
6	469.5	26.1	464	16 Q7U0X2	Q7U0X2 mycobacteri
7	468.5	26.0	464	16 Q53896	Q53896 mycobacteri
8	460	25.5	382	16 Q9CD67	Q9CD67 mycobacteri
9	460	25.5	452	2 Q9Z5G6	Q9Z5G6 mycobacteri
10	407	22.6	542	16 Q9F8X9	Q9F8X9 streptomyc
11	396	22.0	519	16 Q93J30	Q93J30 streptomyc
12	391	21.7	375	16 Q8DG87	Q8DG87 synecococc
13	382	21.2	394	16 P72780	P72780 synecocyst
14	382	21.2	472	16 Q82FM9	Q82FM9 streptomyc
15	382	21.2	619	16 Q82IL8	Q82IL8 streptomyc
16	379.5	21.1	473	16 Q8FR17	Q8FR17 corynebacte

17	373.5	20.7	500	2 Q9KJN6	Q9KJN6 myxococcus
18	372	20.6	407	16 Q8YTF9	Q8YTF9 anabaena sp
19	366	20.3	525	16 Q89QJ8	Q89QJ8 bradyrhizob
20	364.5	20.2	675	15 Q8G6T3	Q8G6T3 bifidobacte
21	363	20.1	395	15 Q7VSC8	Q7VSC8 prochloroco
22	362	20.1	525	16 Q8PMV4	Q8PMV4 xanthomonas
23	358	19.9	413	16 Q8NS10	Q8NS10 corynebacte
24	355.5	19.7	452	16 P73354	P73354 synecocyst
25	354.5	19.7	490	16 Q8XPT5	Q8XPT5 ralstonia s
26	354	19.6	371	16 Q31388	Q31388 bradyrhizob
27	354	19.6	433	16 Q97GD5	Q97GD5 clostridium
28	353	19.6	326	16 Q8DMV9	Q8DMV9 synecococc
29	349	19.4	447	16 Q8R756	Q8R756 thermocanae
30	347.5	19.3	391	16 Q89915	Q89915 clostridium
31	347.5	19.3	432	16 Q7U8K9	Q7U8K9 synecococc
32	347	19.3	374	16 Q7U495	Q7U495 synecococc
33	347	19.3	465	16 Q92OE6	Q92OE6 rhizobium m
34	347	19.3	525	16 Q8PB56	Q8PB56 xanthomonas
35	346	19.2	389	16 Q8DL28	Q8DL28 synecococc
36	345	19.1	339	16 Q89RP2	Q89RP2 bradyrhizob
37	345	19.1	362	16 Q9S2R5	Q9S2R5 streptomyc
38	344	19.1	511	2 Q8RTX2	Q8RTX2 xanthomonas
39	343	19.0	457	16 Q8ZB58	Q8ZB58 versinia pe
40	343	19.0	463	16 Q8DIR0	Q8DIR0 versinia pe
41	342.5	19.0	528	16 Q89S21	Q89S21 bradyrhizob
42	341.5	19.0	514	16 Q9PBA3	Q9PBA3 xylella fas
43	340.5	18.9	505	16 Q8Y016	Q8Y016 ralstonia s
44	338.5	18.8	629	16 Q7UXF4	Q7UXF4 rhodopirell
45	335.5	18.6	429	16 Q8Y7Z0	Q8Y7Z0 anabaena sp

ALIGNMENTS

RESULT 1

O07175 PRELIMINARY: PRT; 355 AA.

AC O07175;
 DT 01-JUL-1997 (TRENBLrel. 04, Created)
 DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Hypothetical protein (Serine protease, putative).
 GN PEPA OR RV0125 OR MTCI418B.07 OR MT0133.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekoa F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the ENBL/GenBank/DBJ databases.

```

CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; Z96071; CAB09453.1; -.
DR EMBL; AE006925; AAK44357.1; -.
DR PIR; F70983; F70983.
DR TIGR; MT0133; -.
DR TubercList; rv0125; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_Ser_trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR008256; Peptidase_S1B_v8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Hypothetical protein; Serine protease; Protease;
KW Complete proteome.
KW Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF-92 CRC64;

Query Match 99.4%; Score 1792; DB 16; Length 355;
Best Local Similarity 99.4%; Pred. No. 3.5e-87;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAVGLGATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAVGLGATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVG 60
QY 61 PQVNVNINIKLGYNNVAGAGTGVDPNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVVG 120
DB 61 PQVNVNINIKLGYNNVAGAGTGVDPNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVVG 120
QY 121 YDRTDVAVLQIRGAGGLPSAALGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTDVAVLQIRGAGGLPSAALGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDLSLTGAEETLNLQIFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFA 240
DB 181 QASDLSLTGAEETLNLQIFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFA 240
QY 241 IPIGQMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVGSAPAASLGIST 300
DB 241 IPIGQMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVGSAPAASLGIST 300
QY 301 GDVITAVDGAIPNSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAIPNSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3
Q50320 PRELIMINARY; PRT; 361 AA.
ID Q50320;
AC Q50320;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 34kDa protein precursor.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JD88/107;
RX MEDLINE=95005449; PubMed=7921248;
RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterisation of a putative serine protease
expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
RL Microbiology 140:1977-1982(1994).
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; Z23092; CAA80638.1; -.
DR PIR; S47170; S47170.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.

Query Match 99.4%; Score 1792; DB 16; Length 355;
Best Local Similarity 99.4%; Pred. No. 3.5e-87;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAVGLGATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAVGLGATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVG 60
QY 61 PQVNVNINIKLGYNNVAGAGTGVDPNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVVG 120
DB 61 PQVNVNINIKLGYNNVAGAGTGVDPNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVVG 120
QY 121 YDRTDVAVLQIRGAGGLPSAALGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTDVAVLQIRGAGGLPSAALGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDLSLTGAEETLNLQIFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFA 240
DB 181 QASDLSLTGAEETLNLQIFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFA 240
QY 241 IPIGQMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVGSAPAASLGIST 300
DB 241 IPIGQMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVGSAPAASLGIST 300
QY 301 GDVITAVDGAIPNSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAIPNSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2
Q7U2S9 PRELIMINARY; PRT; 355 AA.
ID Q7U2S9;
AC Q7U2S9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable serine protease PEPA (EC 3.4.21.-).
GN PEPA OR MB0130.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglimer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,

```

KW Hydrolase; Protease; Serine protease; Signal.
 FT SIGNAL 1 38 POTENTIAL
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78DF3C411 CRC64;
 Query Match 72.5%; Score 1306; DB 2; Length 361;
 Best Local Similarity 70.4%; Pred. No. 1.6e-61;
 Matches 255; Conservative 41; Mismatches 58; Indels 8; Gaps 3;
 QY 1 MNSRSLRSLWSLLSLAAYCLGL-----ATAPQAAPALSODRFPALPLDPS 53
 DB 1 MSKSHHRSVMSWLVGVTVVGLGLGSLGVLAPASAPGLALDRFADPLAPDPS 60
 QY 54 AMYAQAPOVNIINTKLYNNVAGCTGIVDPNGVLTNNHVIAGTDINAFSGSGQT 113
 DB 61 AMVGQVGPQWNIDTFKYNNAVAGCTGIVDPNGVLTNNHVISGATISAFDVGNGQT 120
 QY 114 YGVVVYGVYDRTQDAVQLRGAGGLPSAAGGGVAVGEPVWAMGNSGGGGTTPRAVGRV 173
 DB 121 YAVDVYGVYDRTQDAVQLRGAGGLPSAAGGGVAVGEPVWAMGNSGGGGTTPRAVGRV 180
 QY 174 VALGQTVQASDSLTAETLGLIQDAIOPDGGSGPVVNGLGQVGMNTAASDNFOLS 233
 DB 181 VALNQSVSATDLTGAQENLGLIQADAPIKDGSGGPNWNSAGQVIGVDTTAATSYKMS 240
 QY 234 QGGQGFAPIGQAMALAGQIRSGGGSPVHIQPTAFGLGVVDNNGNGARVQVVGSA 293
 DB 241 -GGQGFAPIGRAMAVANQIRSGAGSNTVHIQPTAFGLGVVDNNGNGARVQVVGSA 299
 QY 294 ASLIGSTGVITAVDGAIPNSATMADALNGHHPGDVSNVWQTSKGGTTRGNVTLAGSP 353
 DB 300 AAGLAFAGGVITGVDTVPINGATSMTEVLVPHHPGDTIAVHFRSVDGGERTANTILASGP 359
 QY 354 PA 355
 DB 360 PA 361
 RESULT 4
 Q9CCY9 ID Q9CCY9 PRELIMINARY; PRT; 354 AA.
 AC Q9CCY9
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Probable secreted serine protease.
 OS Mycobacterium leprae.
 GN Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrall B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 RC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL583926; CAC32191.1; -;
 DR PIR; A87242; A87242.
 DR Leptoma; ML2659; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.

DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR002554; Peptidase_S1.
 DR InterPro; IPR008256; Peptidase_S1B_V8.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR PRINTS; PR00889; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR PRINTS; PR00839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9EA4A CRC64;
 Query Match 70.6%; Score 1272.5; DB 16; Length 354;
 Best Local Similarity 70.1%; Pred. No. 9e-60;
 Matches 251; Conservative 42; Mismatches 58; Indels 7; Gaps 3;
 QY 1 MNSRRSLRSLWSLLSLAAYCLGLATAQA---APPALSQDRFPALPLDPSAMVA 57
 DB 1 MSRPHERSLRSLWSLVSTLAALGLSLAVVPGSATSPSTLALDRFSNRPLPLNPAAM-- 58
 QY 58 QVAPQVNIINTKLYNNVAGCTGIVDPNGVLTNNHVIAGTDINAFSGSGQTGYVD 117
 DB 59 -VAPQVNIINTKLYNNVAGCTGIVDPNGVLTNNHVIAGTDISAFDVGNGKTYVD 117
 QY 118 VVGVDRTQDAVQLRGAGGLPSAAGGGVAVGEPVWAMGNSGGGGTTPRAVGRVVALG 177
 DB 118 VVGVDRTQDAVQLRGAGGLPSAAGGGVAVGEPVWAMGNSGGGGTTPRAVGRVVALG 177
 QY 178 QTVQASDSLTAETLGLIQDAIOPDGGSGPVVNGLGQVGMNTAASDNFOLSOGQ 237
 DB 178 QTVQASDELTGAQETLSGLIQDAIOPDGGSGPVVNSRGQVGMNTAATDNYKM-LGGQ 236
 QY 238 GFAIPIGQAMALAGQIRSGGGSPVHIQPTAFGLGVVDNNGNGARVQVVGSAASLG 297
 DB 237 GFAIPIGQAMVGVGAIIRSGAGSNTVHIQPTAFGLGVVDNNGNGARVQVVGSAASLG 296
 QY 298 ISTGDVITAVDGAIPNSATMADALNGHHPGDVSNVWQTSKGGTTRGNVTLAGSP 355
 DB 297 ISVGDIITSVDGVPISEATAMTNVLVPHHPGETVAVNYSAGGGDLTANVTLAGSP 354
 RESULT 5
 Q8VKA4 ID Q8VKA4 PRELIMINARY; PRT; 446 AA.
 AC Q8VKA4
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Heat shock protein HtrA, putative.
 GN MT1011.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleisichman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF006985; AAK45259.1; -;
 DR TIGR; MT1011; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.

DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 SQ SEQUENCE 446 AA; 44484 MW; 54170CBEA8E872B CRC64;

Query Match 26.1%; Score 469.5; DB 16; Length 446;
 Best Local Similarity 36.0%; Pred. No. 2.9e-17;
 Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;

QY 19 AAVGLGATAPAAQAPALPQSDRFAFPALPLPSA---MVAQVAPVNNIKLGYNNA 75
 DB 102 AASLVGNRAPAGSGGPGVAASAPSIIPANMPGSGVEQAAKVPVSVNLETDLGRQSE 161
 QY 76 VAGTGVIVDPNGVLTNNHVIAGAT-----DINAFVSGQTYGVVDVWGYDRTQDV 127
 DB 162 --EGSGILSAEGLILNNHVIAAAKPPLGSPPKTTVTFSDGRTAPFTVVGADPTSDI 219
 QY 128 AVQLRGAGGLPSAAICGG--VAVGEPVVMGNSGGGGTTPRAVGRVVALGQTVQASDS 185
 DB 220 AVRVQGVSGLTPIISLGSSDLRVGPVLAIGSPFLGLEGT--VTTGIVSALNRPVSTTGE 277
 QY 186 LTGAETLNGLIQFDAAIQDGSQGVVNGLGQVVGWNTA---ASDNFQLSQG--GQGF 239
 DB 278 -AGNQNTVLDIQTDAINFNGSGGALVNNVAQLVGNVSAIATLGADSADAQSGSGLGPF 336
 QY 240 AIPGQAMAIAGQIRSGGSGFTVHIGPTAFGLGLGV--DNNNGARVQVRVGSAPASLGI 298
 DB 337 AIPVDQAKRIADELISTGKA-----SHASLGQVNTDKDTPGAKIVEVWAGGAANAGV 390
 QY 299 STGDVITAVDGPAPNSATAMADALNGHHPGQDVISVNMQTKSGGTRTGNVTLAE 351
 DB 391 PKGVVVTKVDDRPINSADALVAAVRSKAPGATVALTFQDPSGGSRVQVTLGK 443

RESULT 6
 Q7U0X2
 ID Q7U0X2 PRELIMINARY; PRT; 464 AA.
 AC Q7U0X2;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Probable serine protease (Serine proteinase) (EC 3.4.21.-).
 GN MB1009.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=AF2122/97;
 RC MEDLINE=2270107; PubMed=12788972;
 RX GARLIER T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 RL EMBL; BX248337; CAP93870.1.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 464 AA; 46436 MW; AE93A4B3PFA9B3 CRC64;

Query Match 26.1%; Score 469.5; DB 16; Length 446;
 Best Local Similarity 36.0%; Pred. No. 2.9e-17;
 Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;

QY 19 AAVGLGATAPAAQAPALPQSDRFAFPALPLPSA---MVAQVAPVNNIKLGYNNA 75
 DB 120 AASLVGNRAPAGSGGPGVAASAPSIIPANMPGSGVEQAAKVPVSVNLETDLGRQSE 179

QY 76 VAGTGVIVDPNGVLTNNHVIAGAT-----DINAFVSGQTYGVVDVWGYDRTQDV 127
 DB 180 --EGSGILSAEGLILNNHVIAAAKPPLGSPPKTTVTFSDGRTAPFTVVGADPTSDI 237
 QY 128 AVQLRGAGGLPSAAICGG--VAVGEPVVMGNSGGGGTTPRAVGRVVALGQTVQASDS 185
 DB 238 AVRVQGVSGLTPIISLGSSDLRVGPVLAIGSPFLGLEGT--VTTGIVSALNRPVSTTGE 295
 QY 186 LTGAETLNGLIQFDAAIQDGSQGVVNGLGQVVGWNTA---ASDNFQLSQG--GQGF 239
 DB 296 -AGNQNTVLDIQTDAINFNGSGGALVNNVAQLVGNVSAIATLGADSADAQSGSGLGPF 354
 QY 240 AIPGQAMAIAGQIRSGGSGFTVHIGPTAFGLGLGV--DNNNGARVQVRVGSAPASLGI 298
 DB 355 AIPVDQAKRIADELISTGKA-----SHASLGQVNTDKDTPGAKIVEVWAGGAANAGV 408
 QY 299 STGDVITAVDGPAPNSATAMADALNGHHPGQDVISVNMQTKSGGTRTGNVTLAE 351
 DB 409 PKGVVVTKVDDRPINSADALVAAVRSKAPGATVALTFQDPSGGSRVQVTLGK 461

RESULT 7
 OS3896
 ID OS3896 PRELIMINARY; PRT; 464 AA.
 AC OS3896;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative serine protease.
 GN RV0983 OR MTW044.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=H37RV;
 RC MEDLINE=98295987; PubMed=9634230;
 RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RL -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC EMBL; AL021999; CAA17582.1; -.
 DR PIR; C70821; C70821.
 DR Tuberculin; RV0983;
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 464 AA; 46452 MW; AE93BFC53E1EC6F CRC64;

Query Match 26.0%; Score 468.5; DB 16; Length 464;
 Best Local Similarity 36.0%; Pred. No. 3.3e-17;
 Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;

QY 19 AAVGLGATAPAAQAPALPSQDRFADFPALPLDPSA--MVAQVAPQVNNINTKLYNNA 75
Db 120 AASLVGFNAPAGPGGPFVAAARSIPAANNPGSVQVAAKVVPVNNLETDLGRQSE 179
QY 76 VGAGTGIVDPNGVVLNNHVIAGAT-----DINAFSVSGQTYGVVYGYDRTOQV 127
Db 180 --EGSIIILSAGELILTNHVIATAAAPPGLGPPPKTTVTFSDGRTPAPFTVVGADPTSDI 237
QY 128 AVLQIRGAGLPSAALGGG--VAVGEPVVMGNSGGQGTPEAVGRVVALGQTVQASDS 185
Db 238 AVVRVGVSLTFPISLGSSDLRVQPLAIGSPGLGEGT--VTTGIVSALNRPVSTTGE 295
QY 186 LTGAETLNGLIQFDAAIOPGDSGGPVVNGLGQVVGMMNTA-----ASDNFQLSQG--GGGF 239
Db 296 -AGNQNTVLDAIQTDAAINPNSGGALVNMNQLVGVNSAIATLGADSDADAQSGSIGLGF 354
QY 240 AIPIGQAMAIAGQIRSGGSPVTHIGTAPFLGLGVV--DNNNGARVQRVVGSAASLGI 298
Db 355 AIPVDQAKRIADELISTGKA-----SHASLGQVQNTXDTLGAKIVEVAVGGAANAAGV 408
QY 299 STGDVITAVDGAIPNSATAMADALNGHHPGDVSVNMWTKSGTGTGNVTLAE 351
Db 409 PKGVVTVKDDRPINSADALVAARSKAPGATVATFQDPSSGSRIVQVTLGK 461

RESULT 8
Q9CD67
ID Q9CD67 PRELIMINARY; PRT; 382 AA.
AC Q9CD67
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Possible secreted serine protease.
GN ML0176.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churche C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simen S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus";
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AL583917; CAC29684.1; -
DR FIR; H86930; H86930.
DR Lepnoma; ML0176; -
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_1.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 382 AA; 37084 MW; 3DD6DD8AE32A60D CRC64;

Query Match 25.5%; Score 460; DB 16; Length 382;

Best Local Similarity 36.4%; Pred. No. 7.3e-17;
Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;
QY 22 GLGATAPAAQAPALPSQDRFADFPALPLDPSAQAQA---PQVNNINTKLYNNA 77
Db 50 GAGVTPGAASVPAANM-----PSGVEQVAVKVVPVNNLETDLGRQSE-- 94
QY 78 AGTGIVDPNGVVLNNHVIAGAT-----TDINAFSVSGQTYGVVYGYDR 123
Db 95 EGSIIILSAGELILTNHVIATAAAPPGLGPPPKTTVTF--DRTASFTVVGADP 151
QY 124 TDVAVQLRGAGLPSAALGGG--VAVGEPVVMGNSGGQGTPEAVGRVVALGQTVQ 181
Db 152 TSDIADVVRVOSISGLTPTITMGSSADLRVQPVVAVGSPGLAGT--VTSGLVSAALNRPVS 209
QY 182 ASDSLTGAEETLNGLIQFDAAIOPGDSGGPVVNGLGQVVGMMNTA-----SDNFQLSQG 235
Db 210 TTGE-SGNQNTVLDAIQTDAAINPNSGGALVNMNQLVGVNSAIATLGADSDAQS 268
QY 236 GQGFAPIGQAMAIAGQIRSGGSPVTHIGTAPFLGLGVV--DNNNGARVQRVVGSA 294
Db 269 GLGFAIPVDQAKRIADELISTG--KATH-----ASLGQVQVATDKGTPGAKVMDVV 322
QY 295 SLGISTGDVITAVDGAIPNSATAMADALNGHHPGDVSVNMWTKSGTGTGNVTLAE 351
Db 323 NAAVPGVVLTKVDDRLISSADALVAARSKAPGDKVSLTYQDSGSRIVQVTLGK 379

RESULT 9
Q9Z5G6
ID Q9Z5G6 PRELIMINARY; PRT; 452 AA.
AC Q9Z5G6
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative serine protease.
GN MLCB373.28.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris D., Taylor K.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eigmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae";
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL; AL035500; CAB36690.1; -
DR FIR; T45448; T45448.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_1.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 452 AA; 45130 MW; 6CA675EB0911F983 CRC64;

Query Match 25.5%; Score 460; DB 2; Length 452;
 Best Local Similarity 36.4%; Pred. No. 9e-17;
 Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAQAAPPALSSQDRFADFPALPLDPSAMVAQV----PQVNNITKLGYNNAV 77
 DB 120 GAGPVTGPAASVPAAN-----PSGSVEQVAVKVPVSMLETLGRQSE-- 164

QY 78 ACTGIVDPNGVLTNNHVIAGA-----TDINAFVSGQTYGVVDVGVYDR 123
 DB 165 EGSVILSADGLIILTNHVVAAKAPGCGGGLSPKTTVTF--DGRFASPTVVVGADP 221

QY 124 TDVAVLQLRGAGGLPSAAIIGG--VAVGPPVWAMGSGGQGTTPRAVGRVVALGQTVQ 181
 DB 222 TSDIAVVRVQISGLTPTWSSADLRVGQPVAVGSPJLAGT--VTSGIVSALNRPV 279

QY 182 ASDSLTGAEETLGLIQFDAIOPDGSQGVVNGLGQVGMNTA-----SDNFOLSQG 235
 DB 280 TTGE-SGNQNTVLDATDAINPNSGGALVNMGGQVGVNSAIALGADSGDAQSGSI 338

QY 236 GQGFAPIGQAMATAGQIRSGGSPVHIGPTAFLGLGVVDNNGN-CARQVRVVGAPAA 294
 DB 339 GLGFAIPVDQAKRIADLEITG--KATH-----ASLGQVATDGTGTPKQVMDVVAGAAA 392

QY 295 SLGISTGDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAE 351
 DB 393 NAAVPKGVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDSGSSRTVQVTLCK 449

RESULT 10

Q9FBK9 PRELIMINARY; PRT; 542 AA.
 AC Q9FBK9;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC05149 OR SCP8.12.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Denapaita D., Eichner A., Cullum J.,
 RX Redebach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders S., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL939122; CAC01350.1; -;
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0008508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;

Query Match 22.6%; Score 407; DB 16; Length 542;
 Best Local Similarity 32.6%; Pred. No. 7e-14;
 Matches 116; Conservative 56; Mismatches 138; Indels 46; Gaps 12;

QY 22 GLGLATAPAAQAAPPALSSQDRFADFPALPLDPSAMVAQVPOVNVNTKLGYNNAVAGAGTG 81
 DB 206 GVGTVELP-QAGPEAERD-----PDSVAGIAARALPSVTLH--VSGSEAAGTGTG 254

QY 82 IVIDPNGVLTNNHVI--AGATDINAFVSGQTYGVVDVGVYDRDQVAVLQLRGAGCLP 139
 DB 255 FVLDGRGHILTNHVEFPAGSGEITVTFSNGDTAAEAEVWGRDSGYLAVVKGVTGLT 314

QY 140 SAAITGG--GVAAGBPVWAMGSGGQGTTPRAVGRVVALGQTVQ--SDSLTGABETLNG 196
 DB 315 PMLGNSDNVVRGDPVVAIGAPFLAGT--VTSGIIIAKERPIAGGEGGSDISYVDA 372

QY 197 IQFDAIOPDGSQGVVNGLGQVGMNTA-----ASDNFOLSQGQGFAPICQAM 247
 DB 373 LQTDAPINPGNSGGFLLDARGAIGNSAIRSADSGSTESDDGQAGSIGLGFALPINOQK 432

QY 248 AIAGQIRSGGSPVHIGPTAFLGLGV-VNHN--GNGARVQ-----RVVGSAPASL 296
 DB 433 RVABEL-----INTGKAHPVIGITLDMNYTGDGARVSAGKGGDGPVATTGGPGAKA 484

QY 297 GISTGDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAE 352
 DB 485 GIKFGDVITAVDQGVHSGEELIVKTRHPRGDRLETLQRDGKETKVSILVLSGSG 540

RESULT 11

Q93J30 PRELIMINARY; PRT; 519 AA.
 AC Q93J30;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC03977 OR SCBAC25E3.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Collins M., Harris D.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

[3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:777-96(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939118; CAC44701.1; --
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00555; PDZ; 1.
DR PRINTS; PR0089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR Hydrolase; Protease; Serine protease; Complete proteome.
KW SEQUENCE 519 AA; 50327 MW; 5CB8D5F0CC19E428 CRC64;
Query Match 22.0%; Score 396; DB 16; Length 519;
Best Local Similarity 32.2%; Pred. No. 2.Se-13;
Matches 119; Conservative 54; Mismatches 148; Indels 48; Gaps 13;
QY 14 LLSVLAAGVLG-----LATAPAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
Db 159 LIALVAGGLGGGLGYTLARNDESGTTSASDTGGSVXEDAGTVA-----GVAAKAL 212
QY 61 PQVYNINIKLYNNVAGAGTGIVIDPNGVLTNNHVIAGATDINAFSV--GSGQTYGVVDV 118
Db 213 PSTVTIQAE--GSGEGGCTGTGFVDFKEGHIWNNHVVAAVDDGKLSATFPNGKKYDAEV 271
QY 119 VGYDRTQDVAVLQLRGA--GGLPSAIIQGG--VAGFEPVAMGNSGGGGTTPRAVGRVVA 175
Db 272 VGHAAQGYDVAVIKLENAPSLKPLALGDSKVAVGDSSTIAIGAPFGLSNT--VTGTLISA 329
QY 176 LGQTVQASDSLTAETLNGLIQPDAAIQPDSGGPVVNGIGOVVGNMNTA----ASDNFQ 231
Db 330 KNPVASSDGSADSKASYMSALQTDASINFGNSGGPLLDAGQNVIGINSALQSTNGGFG 389
QY 232 LSQGGQ---GPAIPGQMAIAGQIRSGGSGPTVHIQPTAFPLGLGVVDNNGNARV--QR 286
Db 390 TQAGSICGLGPAIPVNAKFAVQQLIKSGKFPVAKIGASVSL-----EETNGAKLTEQG 444
QY 287 VVGS-----APASLGIETGDTITAVDGAPINATAMADALNGHHPGDVISVNWQTKSG 340
Db 445 VGGSDPVEKGGPADAGLKPQGDVITKLDDRIDVDSGPTLIGEIWTHKFGDEVTVTYE--RGG 503
QY 341 GTRTGNVTL 349
Db 504 KQHTAEVTL 512

RESULT 12
QBDG87 PRELIMINARY; PRT; 375 AA.
AC QBDG87;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Serine protease.
GN TLR2436.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriquchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005377; BAC09988.1; --
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR0089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR Protease; Complete proteome.
KW SEQUENCE 375 AA; 39724 MW; 0D2602EB534915F5 CRC64;
Query Match 21.7%; Score 391; DB 16; Length 375;
Best Local Similarity 33.6%; Pred. No. 3.1e-13;
Matches 108; Conservative 49; Mismatches 114; Indels 50; Gaps 10;
QY 35 PALSQDREAD--FPALPLDPSAMVAQVAVPQVNVNINIKLYNNVAGAGTGIVIDPNGVLTN 93
Db 67 PULSDPFFRQPFGLALPQE-----DRLRGQSGFIIDPSGIVMTN 108
QY 94 NHIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRG--AGGLPSAIIQGG--VAVG 150
Db 109 AHVVSQADTVNV--RLKDGKRVFEGVGVDEVSDLAIVKLKGVTEPLTAPLGDSEVRVG 167
QY 151 EPVWAMGNSGGGGTTPRAVGRVVAIG--QTVQASDSLTAETLNGLIQPDAAIQGDS 208
Db 168 DWAIVAGNPLGLDNT-----VTLGIISTLHRSQAQVIGIPDKLDFIQTDAINPGNS 219
QY 209 GGPVNVNGIQVVGMMNTAASDNFQLSGGGGPAIPGQMAIAGQIRSGGSGSTVHIQ--P 266
Db 220 GGPLLNEAGEVIGINTA-----IRADAMGIGFAIPINKAKALQARLIRGEKIOHAYIGIQM 275
QY 267 TAFPLGLGVVDNNG-----NGARVQVVGSAASLGIETGDTITAVDGAPINSA 315
Db 276 TTTTPAMAKENNANPNPVLPEVNGVLVQLVLPNTPAKAGLRWGDVITAVDGEPIISA 335
QY 316 TAMADALNGHHPGDVISVNWQ 336
Db 336 DQLQITIVDSAAVGVNLITVQ 356
RESULT 13
P72780

Query Match	21.2%	Score 382;	DB 16;	Length 619;
Best Local Similarity	33.7%;	Pred. No. 1.7e-12;		
Matches 114;	Conservative 53;	Mismatches 127;	Indels 44;	Gaps 12;

Qy	38	SQRFDAPALPLDPSAMVAQVQVNINTKLGYNNAVGAGTGIVDPNGVLVTNNHVI	97
	:::::	:::::	:::::
Dd	298	SEERAADSA-----GIAARALFSVTLHVK--GSAAEGTGTFVLDGRGHILTNHHVV	349
	:::::	:::::	:::::
Qy	98	--AGATINAFSVSGSGTYGVDDVVGYPRTQDAVLQLRGAGGLSPAAIGG--GVAVEGPV	153
	:::::	:::::	:::::
Dd	350	EPAGSSEISVTFSSGGTAKATAVVRDSGYDLNAVKYSGVGLXPMLNSDNVOVGDPV	409
	:::::	:::::	:::::
Qy	154	VAMNMSGQGQTTPRAVFPFRVVALGQTVOA--SDSITGAETLNLGIQFDAAIQPDSGGPV	212
	:::::	:::::	:::::
Dd	410	VAICAPRLANT--VTSCTISAKRPPTTAGCEKGDSYVDALOTDAPTNPNGSGGPL	467
	:::::	:::::	:::::

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 40.5137 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MEHHHHAPPALSQDRFADF.....QTKSGGTRTGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1698	100.0	330	5 AAE29702	Aae29702 Mycobacte
2	1698	100.0	330	5 AAE17566	Aae17566 Mycobacte
3	1695	99.8	330	5 AAE29703	Aae29703 Mycobacte
4	1695	99.8	330	5 AAE17567	Aae17567 Mycobacte
5	1695	99.8	330	7 ADA26372	Ada26372 Mycobacte
6	1695	99.8	723	7 ADA26354	Ada26354 Mycobacte
7	1695	99.8	1010	7 ADA26356	Ada26356 Mycobacte
8	1691	99.6	330	7 ADA26371	Ada26371 Mycobacte
9	1645	96.9	355	2 AAY05000	Aay05000 Mycobacte
10	1645	96.9	355	4 AAG22137	Aag22137 Mycobacte
11	1645	96.9	355	4 AAG81110	Aag81110 Mycobacte
12	1645	96.9	379	2 AAY04830	Aay04830 Mycobacte
13	1634	96.2	355	2 AAW32367	Aaw32367 Mycobacte
14	1634	96.2	355	2 AAW32435	Aaw32435 Mycobacte
15	1634	96.2	355	2 AAW44307	Aaw44307 Mycobacte
16	1634	96.2	355	2 AAW81670	Aaw81670 M. tuberc
17	1634	96.2	355	2 AAY38972	Aay38972 M. tuberc
18	1634	96.2	355	2 AAY39109	Aay39109 M. tuberc
19	1634	96.2	355	4 AAU01890	Aau01890 M. tuberc
20	1634	96.2	355	5 AAE29701	Aae29701 Mycobacte
21	1634	96.2	355	5 AAE17565	Aae17565 Mycobacte
22	1003.5	59.1	1016	7 ADA26370	Ada26370 M. bovis
23	995	58.6	1154	7 ADA26368	Ada26368 Mycobacte
24	994.5	58.6	825	7 ADA26366	Ada26366 Mycobacte
25	993.5	58.5	813	7 ADA26367	Ada26367 Mycobacte

26	992	58.4	1022	7 ADA26369	Ada26369 Mycobacte
27	987.5	58.2	875	7 ADA26365	Ada26365 Mycobacte
28	987	58.1	596	2 AAY32070	Aay32070 Mycobacte
29	987	58.1	596	5 AAE29710	Aae29710 Mycobacte
30	987	58.1	596	5 AAE17574	Aae17574 Mycobacte
31	987	58.1	599	5 AAU74599	Aau74599 Antigenic
32	987	58.1	729	4 AAO22142	Aao22142 Ra12-H9-3
33	987	58.1	729	5 AAE29708	Aae29708 Mycobacte
34	987	58.1	729	5 AAE17572	Aae17572 Mycobacte
35	987	58.1	729	7 ADA26373	Ada26373 Mycobacte
36	987	58.1	930	5 AAE29731	Aae29731 Mycobacte
37	987	58.1	930	7 ADA26364	Ada26364 Mycobacte
38	984	58.0	729	5 AAE29709	Aae29709 Mycobacte
39	984	58.0	729	5 AAE17573	Aae17573 Mycobacte
40	984	58.0	729	7 ADA26374	Ada26374 Mycobacte
41	982	57.8	726	5 AAU74588	Aau74588 Antigenic
42	982	57.8	729	2 AAY32059	Aay32059 Mycobacte
43	981	57.8	195	5 AAE29704	Aae29704 Mycobacte
44	981	57.8	195	5 AAE17568	Aae17568 Mycobacte
45	676	39.8	231	2 AAY32071	Aay32071 Mycobacte

ALIGNMENTS

RESULT 1
AAE29702
ID AAE29702 standard; protein; 330 AA.
XX
AC AAE29702;
XX
DI 27-JAN-2003 (first entry)
XX
DE Mycobacterium tuberculosis mature Ra35 antigenic protein.
XX
KW Vaccine; immunity; diagnostic agent; gene therapy; Ra35 antigen.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 182
FT /note= "Encoded by GAG"
FT Misc-difference 183
FT /note= "Encoded by GCG"
XX
PN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WPI: 2002-759844/82.
XX
DR N-PSDB; ADA47077.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeifP, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.
XX
PS Disclosure; Page 79-80; 155pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected from TSA, LeifP, M15, and 6H polynucleotides. Sequences of the invention

CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis mature Ra35 (N-terminus of MTB32A; Ra35FL) antigenic protein
 XX
 SQ Sequence 330 AA;

Query Match 100.0%; Score 1698; DB 5; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1.9e-122;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSDQRFADFPALDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGIVID 60
 Db 1 MHHHHHAPPALSDQRFADFPALDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGIVID 60
 QY 61 PNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 Db 61 PNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAAIQP 180
 Db 121 GVAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAAIQP 180
 QY 181 GDSGPPVNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 Db 181 GDSGPPVNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVNDNNGARVQVVGSAAPASLGISTGDTVITAVDGAIPNSATAMADALNGH 300
 Db 241 PTAFLGLGVNDNNGARVQVVGSAAPASLGISTGDTVITAVDGAIPNSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 2

AAE17566

ID AAE17566 standard; protein; 330 AA.

XX

AC AAE17566;

XX

DT 22-APR-2002 (first entry)

XX

DE Mycobacterium sp. MTB32A (Ra35FL) mature protein.

XX

KW Fusion protein; antigen; serological sensitivity; immune response;

XX

KW tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.

XX

OS Mycobacterium sp.

XX

PH Key Location/Qualifiers

FT Region

FT 8..202

FT /note= "Ra35 N-terminal peptide"

FT Misc-difference 182

FT /note= "Encoded by GAG"

FT Misc-difference 183

FT /note= "Encoded by GCG"

FT Region

FT 199..330

FT /note= "Ra35 C-terminal peptide, Ra12"

XX

FN WO200198460-A2.

XX

PD 27-DEC-2001.

XX

PF 20-JUN-2001; 2001WO-US019959.

XX

PR 20-JUN-2000; 2000US-00597796.

XX

PR 01-FEB-2001; 2001US-0265737P.

XX

PA (CORI-) CORIXA CORP.
 XX Skeiky Y, Reed S, Alderson M;
 PI WPI; 2002-147798/19.
 DR N-PSDB; AAD28336.
 DR
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 70; Fig 6; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human, and
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) mature protein
 XX
 SQ Sequence 330 AA;

Query Match 100.0%; Score 1698; DB 5; Length 330;

Best Local Similarity 100.0%; Pred. No. 1.9e-122;

Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGIVID 60
 Db 1 MHHHHHAPPALSDQRFADFPALDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGIVID 60
 QY 61 PNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 Db 61 PNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAAIQP 180
 Db 121 GVAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAAIQP 180
 QY 181 GDSGPPVNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 Db 181 GDSGPPVNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVNDNNGARVQVVGSAAPASLGISTGDTVITAVDGAIPNSATAMADALNGH 300
 Db 241 PTAFLGLGVNDNNGARVQVVGSAAPASLGISTGDTVITAVDGAIPNSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 3

AAE29703

ID AAE29703 standard; protein; 330 AA.

XX

AC AAE29703;

XX

DT 27-JAN-2003 (first entry)

XX

DE Mycobacterium tuberculosis Ra35FLMutSA mutant antigenic protein.
 XX Vaccine; immunity; diagnostic agent; gene therapy; Ra35FLMutSA; mutant;
 KW antigen; mutein.
 XX Mycobacterium tuberculosis.
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 183 /note= "Wild type Ser substituted with Ala"
 FT WT
 XX WO200272792-A2.
 XX
 XX 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US008223.
 XX
 PR 13-MAR-2001; 2001US-0275837P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Brannon M, Guderian J;
 XX
 XX WPI; 2002-759844/82.
 DR N-PSDB; AAD47078.
 XX
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX
 XX Disclosure; Page 81-82; 155pp; English.
 PS
 XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis Ra35FLMutSA mutant antigenic protein
 XX
 XX Sequence 330 AA;

Query Match 99.8%; Score 1695; DB 5; Length 330;
 Best Local Similarity 99.7%; Pred. No. 3.2e-122;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHFHHHHPALSGDRPADFPALPLDPSAVQGVQVNVNKLGVNNAVGAGTGIVD 60
 Db 1 MHFHHHHPALSGDRPADFPALPLDPSAVQGVQVNVNKLGVNNAVGAGTGIVD 60
 Qy 61 PNGVLTNNHVIAGTDINAFSGGQTYGVVYVGYDRTQDVAVLQLRGAGGLPSAAIGG 120
 Db 61 PNGVLTNNHVIAGTDINAFSGGQTYGVVYVGYDRTQDVAVLQLRGAGGLPSAAIGG 120
 Qy 121 GVAVGFPVWAMNGSGGGTTPRAVPGRVVWALGTQVQASDLTCAETLNGLIQFDAAIQP 180
 Db 121 GVAVGFPVWAMNGSGGGTTPRAVPGRVVWALGTQVQASDLTCAETLNGLIQFDAAIQP 180
 Qy 181 GDSGGPWNGLGVQVGNNTAASDNFQISQGGQGFAPIGQMAIAGQIRSGGSPVTHIG 240
 Db 181 GDAGGPVWNGLVQVGNNTAASDNFQISQGGQGFAPIGQMAIAGQIRSGGSPVTHIG 240
 Qy 241 PTAFLGLGVVDNNGNGARVQRVVGSPAASLGISTGDTVITAVDGPINSATAMADALNGH 300
 Db 241 PTAFLGLGVVDNNGNGARVQRVVGSPAASLGISTGDTVITAVDGPINSATAMADALNGH 300

Qy 301 HPGDVISVTWQTKSGGTGTGNVTLAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTGTGNVTLAEGPPA 330
 RESULT 4
 AAEL17567
 ID AAEL17567 standard; protein; 330 AA.
 XX
 AC AAEL17567;
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium species Ra35FL mature protein mutant (S183A), Ra35FLMutSA.
 XX
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; Ra32FLMutSA protein; mutant; mutein.
 XX
 OS Mycobacterium sp.
 XX
 PH Key Location/Qualifiers
 FT Region 8.202
 FT Misc-difference 183 /note= "Ra35 N-terminal peptide"
 FT WT
 XX
 FT /note= "Wild type Ser substituted with Ala"
 FT 199..330
 XX
 FT /note= "Ra35 C-terminal peptide, Ra12"
 FT WT
 XX
 PN WO200198460-A2.
 XX
 XX 27-DEC-2001.
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 XX
 PR 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0285737P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Alderson M;
 XX
 XX WPI; 2002-147798/19.
 DR N-PSDB; AAD29337.
 XX
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 PS Claim 73; Fig 6; 136pp; English.
 XX
 CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC Ra32FL mature protein mutant (S183A), Ra35FLMutSA
 XX
 XX Sequence 330 AA;

Query Match 99.8%; Score 1695; DB 5; Length 330;
 Best Local Similarity 99.7%; Pred. No. 3.2e-122;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGIVID 60
 DB 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGIVID 60

QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 DB 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120

QY 121 GVAVGEPPVAMGNSGGGQTPRAVPGRVVAGLQTVQASDSLTGAEETLNGLIQFDAAIQ 180
 DB 121 GVAVGEPPVAMGNSGGGQTPRAVPGRVVAGLQTVQASDSLTGAEETLNGLIQFDAAIQ 180

QY 181 GDSGGPVVNGLGQVVGNTAASDNFQLSQGGQGFPAIPIGQAMALAGQIRSGGSPVTHIG 240
 DB 181 GDAGGPVVGNGLGQVVGNTAASDNFQLSQGGQGFPAIPIGQAMALAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGDTGVTAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGDTGVTAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 5
 ADA26372
 ID ADA26372 standard; protein; 330 AA.
 XX ADA26372;
 AC ADA26372;
 DT 20-NOV-2003 (first entry)
 DE Mycobacterium MTB32AmuSA protein.
 XX fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine; mutant; mutein.
 XX Synthetic.
 OS Mycobacterium sp.
 XX Key Location/Qualifiers
 FT Misc-difference 183
 FT /note= "Wild-type Ser substituted by Ala"
 XX WO2003070187-A2.
 XX 28-AUG-2003.
 XX 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Guderian J, Reed S;
 XX WPI; 2003-697554/66.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX Disclosure; Fig 19; 112pp; English.
 PS The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and

CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX Sequence 330 AA;

Query Match 99.8%; Score 1695; DB 7; Length 330;
 Best Local Similarity 99.7%; Pred. No. 3.2e-122;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGIVID 60
 DB 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGIVID 60

QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 DB 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120

QY 121 GVAVGEPPVAMGNSGGGQTPRAVPGRVVAGLQTVQASDSLTGAEETLNGLIQFDAAIQ 180
 DB 121 GVAVGEPPVAMGNSGGGQTPRAVPGRVVAGLQTVQASDSLTGAEETLNGLIQFDAAIQ 180

QY 181 GDSGGPVVNGLGQVVGNTAASDNFQLSQGGQGFPAIPIGQAMALAGQIRSGGSPVTHIG 240
 DB 181 GDAGGPVVGNGLGQVVGNTAASDNFQLSQGGQGFPAIPIGQAMALAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGDTGVTAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGDTGVTAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 6
 ADA26354
 ID ADA26354 standard; protein; 723 AA.
 XX ADA26354;
 AC ADA26354;
 DT 20-NOV-2003 (first entry)
 DE Mycobacterium MTB32-MTB39F fusion protein.
 XX fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 XX Chimeric.
 OS Mycobacterium sp.
 XX WO2003070187-A2.
 XX 28-AUG-2003.
 XX 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Guderian J, Reed S;
 XX WPI; 2003-697554/66.
 XX N-PSDB; ADA26353.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.

XX	Claim 4; Fig 2; 11pp; English.
PS	The invention relates to a novel nucleic acid encoding a fusion
XX	polypeptide, comprising a MTB32A and Mtb39 antigen, or MTB32A, and
CC	MTB35A antigen from a Mycobacterium species of the tuberculosis complex.
CC	A polypeptide of the invention has tuberculostatic activity. A
CC	polynucleotide of the invention may have a use in gene therapy, and as a
CC	vaccine. The methods and compositions of the present invention are useful
CC	for diagnosing, preventing and/or treating tuberculosis infection. The
CC	present sequence is used in the exemplification of the invention.
XX	
SQ	Sequence 723 AA;
	Query Match 99.8%; Score 1695; DB 7; Length 723;
	Best Local Similarity 99.7%; Pred. No. 8.6e-122;
	Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MHHHHHHAPPALSODRPAFPALPLDPSAMVAQVGPQVVNINTKLGYNNAVAGCTGVID 60
Dd	1 MHHHHHHAPPALSODRPAFPALPLDPSAMVAQVGPQVVNINTKLGYNNAVAGCTGVID 60
Qy	61 PNGVLVTNNHVIAGATDINAFSVSGGTGYGVYDVGYDTQDAVLQLRGAGGLPSAAIGG 120
Dd	61 PNGVLVTNNHVIAGATDINAFSVSGGTGYGVYDVGYDTQDAVLQLRGAGGLPSAAIGG 120
Qy	121 GVAYGPEPVVMNSGGGGTTPRAYGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAIQP 180
Dd	121 GVAYGPEPVVMNSGGGGTTPRAYGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAIQP 180
Qy	181 GDSSGGPVGNGLGQVVGMMTAASDNFOLSQGGGFALPIGOAWATAGQIRSGGGSPVTHIG 240
Dd	181 GDAGGPVGNGLGQVVGMMTAASDNFOLSQGGGFALPIGOAWATAGQIRSGGGSPVTHIG 240
Qy	241 PTAFLGLGWNDNGGARVQRVWGSAPASLIGISTGDVITAVDGAPINSATAMADALNGH 300
Dd	241 PTAFLGLGWNDNGGARVQRVWGSAPASLIGISTGDVITAVDGAPINSATAMADALNGH 300
Qy	301 HPGDIVISVTWTKSGGTRTGNVTLAEGPPA 330
Dd	301 HPGDIVISVTWTKSGGTRTGNVTLAEGPPA 330

RESULT 7	
ADA26356	
ID	ADA26356 standard; protein; 1010 AA.
XX	
AC	ADA26356;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Mycobacterium MTB-102F fusion protein.
XX	
XX	fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW	tuberculosis; tuberculostatic; gene therapy; vaccine.
XX	
OS	Chimeric.
OS	Mycobacterium sp.
XX	
FN	W02003070187-A2.
XX	
PD	28-AUG-2003.
XX	
PF	18-FEB-2003; 2003WO-US004903.
XX	
PR	15-FEB-2002; 2002US-0357351P.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Skeiky Y, Guderian J, Reed S;
XX	
DR	WPI; 2003-697554/66.
DR	N-PSDB; ADA26355.

XX	New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT	and MTB39 antigens, with or without the MTB85A antigen, from a
PT	Mycobacterium species, useful for diagnosing, preventing and/or treating
PT	tuberculosis infection.
XX	
PS	Claim 6; Fig 4; 112pp; English.
XX	
CC	The invention relates to a novel nucleic acid encoding a fusion
CC	polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC	MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC	A polypeptide of the invention has tuberculostatic activity. A
CC	polynucleotide of the invention may have a use in gene therapy, and as a
CC	vaccine. The methods and compositions of the present invention are useful
CC	for diagnosing, preventing and/or treating tuberculosis infection. The
CC	present sequence is used in the exemplification of the invention.
XX	
SQ	Sequence 1010 AA;
	Query Match 99.8%; Score 1695; DB 7; Length 1010;
	Best Local Similarity 99.7%; Pred. No. 1.3e-121;
	Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MHHHHHAPPALSOQRADFPALPLDPSAMVAOYGPVWNINTKLGYNNAVGAGTGIVLD 60
Dd	1 MHHHHHHAPPALSODRFADFALPLDPSAMVAOYGPVWNINTKLGYNNAVGAGTGIVLD 60
Qy	61 PNGVLVTNNHVIAGATDINAFSVSGGGTGYVDVVGYDRYTQDVAVLQLRGAGLPAAIIG 120
Dd	61 PNGVLVTNNHVIAGATDINAFSVSGGGTGYVDVVGYDRYTQDVAVLQLRGAGLPAAIIG 120
Qy	121 GVAVGEPPVAMNSGGGGGTTPRVPGRVVALGQTVOASDSLTGAETLNGLIQFDAAIQP 180
Dd	121 GVAVGEPPVAMNSGGGGGTTPRVPGRVVALGQTVOASDSLTGAETLNGLIQFDAAIQP 180
Qy	181 GDSCGPPVVGGLGVVGMNTAASDNFQLSQQGGGFAIPICQAMAIAQIRSGCGSPTVHIG 240
Dd	181 GDAGCPPVVGGLGVVGMNTAASDNFQLSQQGGGFAIPICQAMAIAQIRSGCGSPTVHIG 240
Qy	241 PTAFLGLGVNDNNGARVQVRVGSAPAAISLGISTGDVITAVDGGAPINSATAMADALNGH 300
Dd	241 PTAFLGLGVNDNNGARVQVRVGSAPAAISLGISTGDVITAVDGGAPINSATAMADALNGH 300
Qy	301 HPGDVISVTWQTSKGGTRTGNVTLAEGPPA 330
Dd	301 HPGDVISVTWQTSKGGTRTGNVTLAEGPPA 330
RESULT 8	
ADA26371	
ID	ADA26371 standard; protein; 330 AA.
XX	
AC	ADA26371;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Mycobacterium wild type MBT32A protein.
XX	
KW	fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
XW	tuberculosis; tuberculostatic; gene therapy; vaccine.

RESULT 8	
ADA26371	
ID	ADA26371 standard; protein; 330 AA.
XX	
AC	ADA26371;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Mycobacterium wild type MBT32A protein.
XX	
KW	fusion polypeptide; MTB32A; MTB39; antigen;
XX	tuberculosis; tuberculostatic; gene therapy
XX	
OS	Chimeric.
OS	Mycobacterium sp.
XX	
PN	WO2003070187-A2.
XX	
PD	28-AUG-2003.
XX	
PF	18-FEB-2003; 2003WO-US004903.
XX	
PR	15-FEB-2002; 2002US-0357351P.
XX	
PA	{CORI-} CORIXA CORP.

```

XX Skeiky Y, Guderian J, Reed S;
PI WPI; 2003-697554/66.
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX and MTB39 antigens, with or without the MTB85A antigen, from a
XX Mycobacterium species, useful for diagnosing, preventing and/or treating
XX tuberculosis infection.
XX
XX Disclosure; Fig 19; 112pp; English.
XX
XX The invention relates to a novel nucleic acid encoding a fusion
XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
XX MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A
XX polynucleotide of the invention may have a use in gene therapy, and as a
XX vaccine. The methods and compositions of the present invention, are useful
XX for diagnosing, preventing and/or treating tuberculosis infection. The
XX present sequence is used in the exemplification of the invention.
XX
XX Sequence 330 AA;
SQ
    Query Match          99.6%; Score 1691; DB 7; Length 330;
    Best Local Similarity 99.7%; Pred. No. 6.6e-122;
    Matches 329; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHEHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVID 60
DB 1 MHHEHHHAPPALSQDRFADFPALPLDPSAHVAQVGPQVNNINTKLGYNNAVGAGTGIVID 60

QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAIGG 120
DB 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAIGG 120

QY 121 GVAVGEPVWAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
DB 121 GVAVGEPVWAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180

QY 181 GDSGGPVNGLGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDSGGPVNGLGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVVDDNNGARVORVVGSAAPASLIGISTGDTVITAVDGPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDDNNGARVORVVGSAAPASLIGISTGDTVITAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330

RESULT 9
AAU05000
ID AAU05000 standard; protein; 355 AA.
XX
XX AAU05000;
XX
XX 06-JUL-1999 (first entry)
XX
XX Mycobacterium species protein sequence 50D.
XX
XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX hybridisation; detection; vaccine; immunisation; infection.
XX
XX Mycobacterium sp.
XX
XX WO9909186-A2.
XX
XX 25-FEB-1999.
XX
XX 14-AUG-1998; 98WO-FR001813.
XX

```

```

PR 14-AUG-1997; 97FR-00010404.
XX 11-SEP-1997; 97FR-00011325.
XX
XX (INSP ) INST PASTEUR.
XX
XX Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;
XX Goguet De La Salmoniere Y;
XX
XX WPI; 1999-181045/15.
XX DR N-PSDB; AAX34251.
XX
XX Mycobacterial DNA vectors containing reporter constructs - for
XX identifying coding or promoter sequences involved in infection-associated
XX protein expression.
XX
XX Claim 32; Fig 50D; 309pp; French.
XX
XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
XX from various Mycobacterium species microorganisms. The encoding
XX nucleotide sequences can be used as primers and probes for methods for
XX detecting and identifying mycobacteria, especially belonging to the M.
XX tuberculosis complex. The encoded proteins can be used in vaccines for
XX immunisation against a bacterial or viral infection
XX
XX Sequence 355 AA;
SQ
    Query Match          96.9%; Score 1645; DB 2; Length 355;
    Best Local Similarity 100.0%; Pred. No. 2.5e-118;
    Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLT 92

QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIIGGVAVGEP 152

QY 128 VVAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIOPGDSGGPV 187
DB 153 VVAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIOPGDSGGPV 212

QY 188 VNGLQVVGMMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGTAFGL 247
DB 213 VNGLQVVGMMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGTAFGL 272

QY 248 GVDNNGNGARVORVVGSAAPASLIGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVDNNGNGARVORVVGSAAPASLIGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332

QY 308 VTWOTKSGGTRTGNVTLAEGPPA 330
DB 333 VTWOTKSGGTRTGNVTLAEGPPA 355

RESULT 10
AAO22137
ID AAO22137 standard; protein; 355 AA.
XX
XX AAO22137;
XX
XX 03-OCT-2002 (first entry)
XX
XX Mycobacterium tuberculosis MTB32A protein.
XX
XX Rai2; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
XX vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
XX immunogen; cytokine.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200125401-A2.
XX

```

PD 12-APR-2001.
XX 06-OCT-2000; 2000WO-US027652.
XX 07-OCT-1999; 99US-0159585P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Guderian J;
XX WPI; 2001-266299/27.
XX N-PSDB; AAL40768.
XX Recombinant nucleic acid molecule for producing high yield expression of
PT desired fusion polypeptides, encodes fusion polypeptide comprising
PT Mycobacterium tuberculosis coding sequence and heterologous polypeptide.
XX
PS Disclosure; Fig 1; 39pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide, comprising a polynucleotide sequence of Ra12, a 14
CC kDa C-terminal fragment of serine protease antigen MTB32A of
CC Mycobacterium tuberculosis, and a heterologous polynucleotide sequence.
CC The recombinant fusion nucleic acids and polypeptides are useful for
CC providing stable and high yield expression of fusion polypeptides of both
CC eukaryotic and prokaryotic origin and to encode a protein product for use
CC as an antigen for detecting serum antibodies. The presence of serum
CC antibodies to M. tuberculosis antigens in an individual indicates that
CC the individual is infected with it. The fusion polypeptides are useful as
CC sources of proteins for monitoring binding of serum antibodies to fusion
CC proteins and as an immunogen to induce and/or enhance immune responses.
CC The coding sequences can be ligated with a coding sequence of another
CC molecule such as M. tuberculosis antigens, a cytokine or an adjuvant and
CC can be used in vivo as a DNA vaccine. This sequence represents the
CC Mycobacterium tuberculosis MTB32A protein
XX
SQ Sequence 355 AA;

Query Match 96.9%; Score 1645; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.5e-118; Indels 0; Gaps 0;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPNGVWLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPNGVWLT 92

QY 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAGGVAVGEP 152

QY 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPGDSGGPV 212

QY 188 VNLGQVVGWNTAAADNFQISQGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFIQL 247
DB 213 VNLGQVVGWNTAAADNFQISQGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFIQL 272

QY 248 GVVDNNGNGARVORVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGNGARVORVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 332

QY 308 VTWQTKSGGTRTGNVTIAEGPPA 330
DB 333 VTWQTKSGGTRTGNVTIAEGPPA 355

RESULT 11
AAG81110
ID AAG81110 standard; protein; 355 AA.
XX
AC AAG81110;
XX

DT 04-SEP-2001 (first entry)
XX Mycobacterium tuberculosis potential drug target protein SEQ ID 161.
XX Drug target; growth; organism viability; characterisation.
XX Mycobacterium tuberculosis.
XX WO200135317-A1.
XX 17-MAY-2001.
XX 13-NOV-2000; 2000WO-US031152.
XX 12-NOV-1999; 99US-0165086P.
PR 12-NOV-1999; 99US-0165124P.
PR 01-FEB-2000; 2000US-0179531P.
XX
PA (REGC) UNIV CALIFORNIA.
XX Eisenberg D, Rotstein SH, Marcotte EM;
XX WPI; 2001-329193/34.
XX N-PSDB; AAH51961.
XX Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the sequences.
XX
PS Disclosure; Page 157; 207pp; English.
XX
XX This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism
XX
SQ Sequence 355 AA;

Query Match 96.9%; Score 1645; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.5e-118; Indels 0; Gaps 0;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPNGVWLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPNGVWLT 92

QY 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAGGVAVGEP 152

QY 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPGDSGGPV 212

QY 188 VNLGQVVGWNTAAADNFQISQGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFIQL 247
DB 213 VNLGQVVGWNTAAADNFQISQGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFIQL 272

QY 248 GVVDNNGNGARVORVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGNGARVORVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 332

QY 308 VTWQTKSGGTRTGNVTIAEGPPA 330
DB 333 VTWQTKSGGTRTGNVTIAEGPPA 355

Dd 333 VTWQKSGGTRTGNVTLAGPPA 355

RESULT 12

AA04830
ID AAY04830 standard; protein; 379 AA.

AC AAY04830;

DT 06-JUL-1999 (first entry)

DE Mycobacterium species protein sequence 50F.

Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
hybridisation; detection; vaccine; immunisation; infection.

OS Mycobacterium sp.

PN WO9909186-A2.

PD 25-FEB-1999.

PF 14-AUG-1998; 98WO-FR001813.

PR 14-AUG-1997; 97FR-00010404.

PS 11-SEP-1997; 97FR-00011325.

PA (INSP) INST PASTEUR.

PI Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;

PI Goguet De La Salmoniere Y;

XX WPI; 1999-181045/15.

DR N-PSDB; AAX34252.

XX Mycobacterial DNA vectors containing reporter constructs - for
identifying coding or promoter sequences involved in infection-associated
protein expression.

PS Claim 32; Fig 50F; 309pp; French.

XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
from various Mycobacterium species microorganisms. The encoding
nucleotide sequences can be used as primers and probes for methods for
detecting and identifying mycobacteria, especially belonging to the M.
tuberculosis complex. The encoded proteins can be used in vaccines for
immunisation against a bacterial or viral infection

XX Sequence 379 AA;

Query Match 96.9%; Score 1645; DB 2; Length 379;
Best Local Similarity 100.0%; Pred. No. 2.7e-118;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGVIDPNCVLT 67

Dd 57 APPALSQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGVIDPNCVLT 116

QY 68 NNHVIAGATDINAFSVGSGQTYGVVGYDRTQDVAVLQLRGAGLPSAAGGVAVGEP 127

Dd 117 NNHVIAGATDINAFSVGSGQTYGVVGYDRTQDVAVLQLRGAGLPSAAGGVAVGEP 176

QY 128 VVAMGNSGGGGTTPRAVPGRVWALGTQVQASDLSLTGAETLNLGIQFDAAIQPGSGGPV 187

Dd 177 VVAMGNSGGGGTTPRAVPGRVWALGTQVQASDLSLTGAETLNLGIQFDAAIQPGSGGPV 236

QY 188 VNLGLGVGVMNTAASDNFOLSCGGGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGL 247

Dd 237 VNLGLGVGVMNTAASDNFOLSCGGGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGL 296

QY 248 GVVDDNNGNGARVQVRVGSAPASLGISTGVTAVDGPINSATAMADALNGHFGDVIS 307

Dd 297 GVVDDNNGNGARVQVRVGSAPASLGISTGVTAVDGPINSATAMADALNGHFGDVIS 356

QY 308 VTWQKSGGTRTGNVTLAGPPA 330

Dd 357 VTWQKSGGTRTGNVTLAGPPA 379

RESULT 13

AAW32367

ID AAW32367 standard; protein; 355 AA.

XX AAW32367;

DT 13-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen TbRa35.

Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
skin testing; M.tuberculosis.

OS Mycobacterium tuberculosis.

PN WO9709429-A2.

PD 13-MAR-1997.

PF 30-AUG-1996; 96WO-US014675.

PR 01-SEP-1995; 95US-00523435.

PR 22-SEP-1995; 95US-00532136.

PR 22-MAR-1996; 96US-00620280.

PR 05-JUN-1996; 96US-00658800.

PR 12-JUL-1996; 96US-00680573.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PI Vedwick TH, Twardzik DR;

XX WPI; 1997-192904/17.

DR N-PSDB; AAT91414.

XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
useful for diagnosis of M. tuberculosis infection.

PS Example 3; Page 124-126; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an
immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
variant differing only in conservative substitutions and/or
modifications). The present sequence represents a M.tuberculosis antigen,
TbRa35. The immunogenic polypeptide can be used to diagnose
M.tuberculosis infection by forming complexes with specific antibodies in
the sample. Fragments of DNA encoding the immunogenic polypeptide can be
used as diagnostic primers or probes and agents that bind to the antigen,
especially monoclonal antibodies or equivalent polyclonal antibodies, are
also used for diagnosis

XX Sequence 355 AA;

Query Match 96.2%; Score 1634; DB 2; Length 355;
Best Local Similarity 99.4%; Pred. No. 1.7e-117;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGVIDPNCVLT 67

Dd 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGVIDPNCVLT 92

QY 68 NNHVIAGATDINAFSVGSGQTYGVVGYDRTQDVAVLQLRGAGLPSAAGGVAVGEP 127

Dd 93 NNHVIAGATDINAFSVGSGQTYGVVGYDRTQDVAVLQLRGAGLPSAAGGVAVGEP 152

QY 128 VVAMGNSGGGGTTPRAVPGRVWALGTQVQASDLSLTGAETLNLGIQFDAAIQPGSGGPV 187

Db 153 VVAMGNSGGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNLGQVVGWMTAASDNFQSGGGFAIPQGMALAGQIRSGGSPVTHIGPTAFGL 247
Db 213 VNLGQVVGWMTAASDNFQSGGGFAIPQGMALAGQIRSGGSPVTHIGPTAFGL 272
QY 248 GVDNNGNGARVQRVVGSAASLGIISTGDTTAVDGPINSATAMADALNGHHPGDVIS 307
Db 273 GVDNNGNGARVQRVVGSAASLGIISTGDTTAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQKSGGTRTGNVTLAEGPPA 355

RESULT 14
AAW32435
ID AAW32435 standard; protein; 355 AA.
XX
AC AAW32435;
XX
DT 08-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbrA35.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709428-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014674.
XX
PR 01-SEP-1995; 95US-00523436.
PR 22-SEP-1995; 95US-00533634.
PR 22-MAR-1996; 96US-00620874.
PR 05-JUN-1996; 96US-00659683.
PR 12-JUL-1996; 96US-00680574.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX
DR WPI; 1997-192903/17.
DR N-PSDB; AAT91477.
XX
XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT useful in vaccines for prevention or treatment of tuberculosis, also for
PT diagnosis.
XX
PS Example 3; Page 114-116; 168pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,
CC TbrA35. The immunogenic protein, and fusion proteins containing one or
CC more of the proteins or one of the proteins plus ESAT-6, are useful in
CC vaccines, preferably when formulated with a non-specific adjuvant, to
CC induce an immune response against M.tuberculosis (for treatment or
CC prevention)
XX
SQ Sequence 355 AA;

Query Match 96.2%; Score 1634; DB 2; Length 355;
Best Local Similarity 99.4%; Pred. No. 1.7e-117;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSDQRFADFPALPDPSAMVAQVGPQVWNTKLGYNNAVGAGTGIVDPNGVLT 67

Db 33 APPALSDQRFADFPALPDPSAMVAQVGPQVWNTKLGYNNAVGAGTGIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVVGXYDRTODVAVLQLRGAGGLPSAAIGGVAYPE 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVVVGXYDRTODVAVLQLRGAGGLPSAAIGGVAYPE 152
QY 128 VVAMGNSGGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Db 153 VVAMGNSGGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNLGQVVGWMTAASDNFQSGGGFAIPQGMALAGQIRSGGSPVTHIGPTAFGL 247
Db 213 VNLGQVVGWMTAASDNFQSGGGFAIPQGMALAGQIRSGGSPVTHIGPTAFGL 272
QY 248 GVDNNGNGARVQRVVGSAASLGIISTGDTTAVDGPINSATAMADALNGHHPGDVIS 307
Db 273 GVDNNGNGARVQRVVGSAASLGIISTGDTTAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQKSGGTRTGNVTLAEGPPA 355

RESULT 15
AAW64307
ID AAW64307 standard; protein; 355 AA.
XX
AC AAW64307;
XX
DT 17-OCT-2003 (revised)
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbrA35.
XX
KW Tuberculosis; infection; diagnosis; antigen; TbrA35.
XX
OS Mycobacterium tuberculosis; strain H37Ra.
XX
PN WO9816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US018214.
XX
PR 11-OCT-1996; 96US-00729622.
PR 13-MAR-1997; 97US-00818111.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ;
XX
DR WPI; 1998-251292/22.
DR N-PSDB; AAV44355.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
PT develop products for the detection of M. tuberculosis infection and
PT diagnosis of tuberculosis.
XX
PS Example 3; Page 115-116; 250pp; English.
XX
CC This polypeptide comprises Mycobacterium tuberculosis soluble antigen
CC TbrA35. It is encoded by a DNA sequence (see AAV44355) isolated from a M.
CC tuberculosis strain H37Ra expression library with rabbit anti-sera raised
CC against M. tuberculosis supernatant. No significant homology was found
CC between TbrA35 and Genbank database sequences. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of a
CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic kits for

CC detecting M. tuberculosis infection in a patient using the above
CC polypeptides, antibodies or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis. (Updated on 17-Oct-2003 to standardise OS
CC field)

XX SQ Sequence 355 AA;

```
Query Match          96.2%; Score 1634; DB 2; Length 355;
Best Local Similarity 99.4%; Pred. No. 1.7e-117;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALFDPDSAMVAQVQPVNINTKLGYNNAVAGTGVIDPNGVLT 67
Db 33 APPALSQDRFADFPALFDPDSAMVAQVQPVNINTKLGYNNAVAGTGVIDPNGVLT 92

QY 68 NNHVIAGATDINAFSVSGQTYGVYDVTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGQTYGVYDVTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152

QY 128 VVAMNSGGQGGTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQFGDSGGPV 187
Db 153 VVAMNSGGQGGTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQFGDSGGPV 212

QY 188 VNLGLQVVGWNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIGTAFGL 247
Db 213 VNLGLQVVGWNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIGTAFGL 272

QY 248 GVDNNGNGARVQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
Db 273 GVDNNGNGARVQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332

QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355
```

Search completed: June 22, 2004, 17:15:24
Job time : 41.5137 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 11.78 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSDRFDFA.....QTKSGGTRGNVTLAGDPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1634	96.2	355	3	US-08-818-112-79 Sequence 79, Appl
2	1634	96.2	355	4	US-08-818-111-80 Sequence 80, Appl
3	1634	96.2	355	4	US-09-056-556-79 Sequence 79, Appl
4	1634	96.2	355	4	US-09-072-596-80 Sequence 80, Appl
5	1634	96.2	355	4	US-09-072-967-79 Sequence 79, Appl
6	987	58.1	596	4	US-09-287-849-26 Sequence 26, Appl
7	987	58.1	729	4	US-09-223-040-2 Sequence 2, Appl
8	987	58.1	729	4	US-09-287-849-28 Sequence 28, Appl
9	676	39.8	231	4	US-09-287-849-28 Sequence 28, Appl
10	670	39.5	132	3	US-08-818-112-66 Sequence 66, Appl
11	670	39.5	132	4	US-08-818-111-67 Sequence 67, Appl
12	670	39.5	132	4	US-09-056-556-66 Sequence 66, Appl
13	670	39.5	132	4	US-09-072-596-67 Sequence 67, Appl
14	670	39.5	132	4	US-09-072-967-66 Sequence 66, Appl
15	670	39.5	132	4	US-09-636-215-819 Sequence 819, Appl
16	670	39.5	132	4	US-09-636-215-848 Sequence 848, Appl
17	670	39.5	132	4	US-09-685-166A-819 Sequence 819, Appl
18	670	39.5	132	4	US-09-685-166A-848 Sequence 848, Appl
19	633	37.3	224	4	US-09-636-215-825 Sequence 825, Appl
20	633	37.3	224	4	US-09-685-166A-825 Sequence 825, Appl
21	633	37.3	267	4	US-09-643-597-352 Sequence 352, Appl
22	633	37.3	267	4	US-09-606-421B-352 Sequence 352, Appl
23	633	37.3	273	4	US-09-736-457-1864 Sequence 1864, Ap
24	633	37.3	229	4	US-09-643-597-354 Sequence 354, Appl
25	633	37.3	229	4	US-09-606-421B-354 Sequence 354, Appl
26	633	37.3	304	4	US-09-636-215-835 Sequence 835, Appl
27	633	37.3	304	4	US-09-685-166A-835 Sequence 835, Appl

US-08-818-112-79
; Sequence 79, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-79
Query Match 96.2%; Score 1634; DB 3; Length 355;
Best Local Similarity 99.4%; Pred. No. 2.6e-131;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Sequence 1863, App
Sequence 852, App
Sequence 852, App
Sequence 349, App
Sequence 349, App
Sequence 333, App
Sequence 333, App
Sequence 196, App
Sequence 196, App
Sequence 196, App
Sequence 353, App
Sequence 353, App
Sequence 337, App
Sequence 337, App
Sequence 309, App
Sequence 309, App
Sequence 325, App
Sequence 325, App

Db 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNNINTKLYNNNAVAGAGTGVIDPNGVULT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVYVGYDRDQDVAVLQLRGAGGLPSAAIIGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVYVGYDRDQDVAVLQLRGAGGLPSAAIIGGVAVGEP 152
QY 128 VVAMGNSGGGQGTTPRAVPRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Db 153 VVAMGNSGGGQGTTPRAVPRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNGLGQVVGWNTAASDNFQLSQGGQGFPAIPGOAMAIAGQIRSGGSPVTHIGPTAFLGL 247
Db 213 VNGLGQVVGWNTAASDNFQLSQGGQGFPAIPGOAMAIAGQIRSGGSPVTHIGPTAFLGL 272
QY 248 GVVDNNGNGARVQVVGSGAPASLGI STGDVITAVDGPINSATAMADALNGHHFGDVIS 307
Db 273 GVVDNNGNGARVQVVGSGAPASLGI STGDVITAVDGPINSATAMADALNGHHFGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4
US-09-072-596-80
; Sequence 80, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-80

Query Match 96.2%; Score 1634; DB 4; Length 355;
Best Local Similarity 99.4%; Pred. No. 2.6e-131;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVAPQVNNINTKLYNNNAVAGAGTGVIDPNGVULT 67
Db 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNNINTKLYNNNAVAGAGTGVIDPNGVULT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVYVGYDRDQDVAVLQLRGAGGLPSAAIIGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVYVGYDRDQDVAVLQLRGAGGLPSAAIIGGVAVGEP 152
QY 128 VVAMGNSGGGQGTTPRAVPRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Db 153 VVAMGNSGGGQGTTPRAVPRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNGLGQVVGWNTAASDNFQLSQGGQGFPAIPGOAMAIAGQIRSGGSPVTHIGPTAFLGL 247
Db 213 VNGLGQVVGWNTAASDNFQLSQGGQGFPAIPGOAMAIAGQIRSGGSPVTHIGPTAFLGL 272
QY 248 GVVDNNGNGARVQVVGSGAPASLGI STGDVITAVDGPINSATAMADALNGHHFGDVIS 307
Db 273 GVVDNNGNGARVQVVGSGAPASLGI STGDVITAVDGPINSATAMADALNGHHFGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 5
US-09-072-967-79
; Sequence 79, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-79

Query Match 96.2%; Score 1634; DB 4; Length 355;
Best Local Similarity 99.4%; Pred. No. 2.6e-131;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 8 APPALSQDRFADFPALPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVDPNGVWLT 67
Db 33 APPALSQDRFADFPALPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVDPNGVWLT 92
Qy 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
Qy 128 VVAMNSGGGGTTPRAVPGRVVAVLQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 153 VVAMNSGGGGTTPRAVPGRVVAVLQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 212
Qy 188 VNLGQVVGMMNTAASDFQISQGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 247
Db 213 VNLGQVVGMMNTAASDFQISQGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 272
Qy 248 GVVDNNGGARVORVVGSAAPASIGSTGVITAVDGAIPINSATAMADALNGHHGPDVIS 307
Db 273 GVVDNNGGARVORVVGSAAPASIGSTGVITAVDGAIPINSATAMADALNGHHGPDVIS 332
Qy 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 6

US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 26
LENGTH: 596
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:

OTHER INFORMATION: Description of Artificial Sequence:bi-fusion

US-09-287-849-26

Query Match 59.1%; Score 987; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 6.1e-76;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 8 APPALSQDRFADFPALPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVDPNGVWLT 67
Db 402 APPALSQDRFADFPALPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVDPNGVWLT 461

Qy 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
Db 462 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 521
Qy 128 VVAMNSGGGGTTPRAVPGRVVAVLQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 522 VVAMNSGGGGTTPRAVPGRVVAVLQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 581
Qy 188 VNLGQVVGMMNTAAS 202
Db 582 VNLGQVVGMMNTAAS 596

RESULT 7

US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6545222

GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009010US
; CURRENT FILING DATE: 1998-12-30
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

OTHER INFORMATION: Description of Artificial Sequence:tri-fusion

US-09-223-040-2

Query Match 58.1%; Score 987; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.9e-76;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 8 APPALSQDRFADFPALPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVDPNGVWLT 67
Db 535 APPALSQDRFADFPALPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVDPNGVWLT 594
Qy 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 654
Qy 128 VVAMNSGGGGTTPRAVPGRVVAVLQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 655 VVAMNSGGGGTTPRAVPGRVVAVLQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 714
Qy 188 VNLGQVVGMMNTAAS 202
Db 715 VNLGQVVGMMNTAAS 729

RESULT 8

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US

```
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

Query Match      58.1%; Score 987; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.9e-76;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 APPALSQDRPADPALPLDPSAMVAQVGPVNVNITKLGYNNAVAGAGTGIVIDPENGVLVT 67
DB      535 APPALSQDRPADPALPLDPSAMVAQVGPVNVNITKLGYNNAVAGAGTGIVIDPENGVLVT 594

QY      68 NNHVIAGATDINAFSGSGGTGYGVVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB      595 NNHVIAGATDINAFSGSGGTGYGVVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 654

QY      128 VVAMGNSGGGGTPRAVPGRVVVALGQTVQASDSLTGABETLNGLIQFDAAIQGDSGGPV 187
DB      655 VVAMGNSGGGGTPRAVPGRVVVALGQTVQASDSLTGABETLNGLIQFDAAIQGDSGGPV 714

QY      188 VVNGLGQVVGNTAAAS 202
DB      715 VVNGLGQVVGNTAAAS 729

RESULT 9
US-09-287-849-28
; Sequence 28, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 231

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-28

Query Match      39.8%; Score 676; DB 4; Length 231;
Best Local Similarity 99.2%; Pred. No. 5.2e-50;
Matches 132; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      198 NTAASDNFQLSQGGGFAPIGQAVAIAGQIRSGGSGPTVHIGTAFGLGVDNNGNGA 257
DB      8 HTAASDNFQLSQGGGFAPIGQAVAIAGQIRSGGSGPTVHIGTAFGLGVDNNGNGA 67

QY      258 RVQRVVGSAAPASLIGSTGCVITAVDGPINSATAMADALNGHHPGDDVISVTWTKSGGT 317
DB      68 RVQRVVGSAAPASLIGSTGCVITAVDGPINSATAMADALNGHHPGDDVISVTWTKSGGT 127

QY      318 RTGNVTLAEGPPA 330
DB      128 RTGNVTLAEGPPA 140

RESULT 10
US-08-818-112-66
; Sequence 66, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-66

Query Match      39.5%; Score 670; DB 3; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 199 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
DB 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
QY 259 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 318
DB 61 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
DB 121 TGNVTLAEGPPA 132

RESULT 11
US-08-818-111-67
; Sequence 67, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-67

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 199 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
DB 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
QY 259 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 318
DB 61 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
DB 121 TGNVTLAEGPPA 132

RESULT 12
US-09-056-556-66
; Sequence 66, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-66

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 199 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
DB 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
QY 259 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 318
DB 61 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
DB 121 TGNVTLAEGPPA 132

RESULT 13
US-09-072-596-67
; Sequence 67, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.

APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 210121.417C9
REFERENCE/DOCKET NUMBER:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 622-4900
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-67
Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 199 TAASDNFQLSQGGQFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 258
Db 1 TAASDNFQLSQGGQFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 60
QY 259 VQRVVGSAAPASLGISTGCVITAVDGPINSATAMADALNGHHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGCVITAVDGPINSATAMADALNGHHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132
RESULT 14
US-09-072-967-66
Sequence 66, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 210121.411C9
REFERENCE/DOCKET NUMBER:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 622-4900
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-66
Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 199 TAASDNFQLSQGGQFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 258
Db 1 TAASDNFQLSQGGQFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 60
QY 259 VQRVVGSAAPASLGISTGCVITAVDGPINSATAMADALNGHHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGCVITAVDGPINSATAMADALNGHHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132
RESULT 15
US-09-636-215-819
Sequence 819, Application US/09636215
Patent No. 6620922
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.42717C17
CURRENT APPLICATION NUMBER: US/09/636,215
CURRENT FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 852
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 819

```
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-636-215-819

Query Match      39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 199 TAASDNFOLSGGGGFAIPICQAWAIAQIIRSGGSPVHIGPTAFILGLGVVDNNGAR 258
      |||||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||
Db 1 TAASDNFOLSGGGGFAIPICQAWAIAQIIRSGGSPVHIGPTAFILGLGVVDNNGAR 60
      |||||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||

Qy 259 VQVVGSAPAASLGIISTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQTKSGGTR 318
      |||||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||
Db 61 VQVVGSAPAASLGIISTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQTKSGGTR 120
      |||||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||

Qy 319 TGNVTLAEGPPA 330
      |||||||U|||||
Db 121 TGNVTLAEGPPA 132
      |||||||U|||||
```

Search completed: June 22, 2004, 17:27:28
Job time : 11.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 30.5644 Seconds

(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSQRDFADF.....QTKSGGTRGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 28213646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1698	100.0	330	12	US-09-886-349A-4
2	1698	100.0	330	14	US-10-098-732A-4
3	1698	100.0	330	15	US-10-369-983-19
4	1695	99.8	330	12	US-09-886-349A-6
5	1695	99.8	330	14	US-10-098-732A-6
6	1695	99.8	330	15	US-10-369-983-20
7	1695	99.8	723	15	US-10-369-983-2
8	1695	99.8	1010	15	US-10-369-983-4
9	1645	96.9	355	9	US-09-712-363-161
10	1634	96.2	355	12	US-09-886-349A-2
11	1634	96.2	355	14	US-10-193-002-80
12	1634	96.2	355	14	US-10-084-843-79
13	1634	96.2	355	14	US-10-098-732A-2
14	1003.5	59.1	1016	15	US-10-369-983-18
15	995	58.6	1154	15	US-10-369-983-16

16	994.5	58.6	825	15	US-10-369-983-14	Sequence 14, Appl
17	993.5	58.5	813	15	US-10-369-983-15	Sequence 15, Appl
18	992	58.4	1022	15	US-10-369-983-17	Sequence 17, Appl
19	987.5	58.2	875	15	US-10-369-983-13	Sequence 13, Appl
20	987	58.1	596	9	US-09-287-849-26	Sequence 26, Appl
21	987	58.1	596	12	US-09-886-349A-20	Sequence 20, Appl
22	987	58.1	596	14	US-10-359-460-26	Sequence 26, Appl
23	987	58.1	596	14	US-10-098-732A-20	Sequence 20, Appl
24	987	58.1	729	9	US-09-287-849-2	Sequence 2, Appl
25	987	58.1	729	12	US-09-886-349A-16	Sequence 16, Appl
26	987	58.1	729	14	US-10-359-460-2	Sequence 2, Appl
27	987	58.1	729	14	US-10-098-732A-16	Sequence 16, Appl
28	987	58.1	729	15	US-10-369-983-21	Sequence 21, Appl
29	987	58.1	729	15	US-10-359-459-2	Sequence 2, Appl
30	987	58.1	930	14	US-10-098-732A-65	Sequence 65, Appl
31	987	58.1	930	15	US-10-369-983-12	Sequence 12, Appl
32	984	58.0	729	12	US-09-886-349A-18	Sequence 18, Appl
33	984	58.0	729	14	US-10-098-732A-18	Sequence 18, Appl
34	984	58.0	729	15	US-10-369-983-22	Sequence 22, Appl
35	981	57.8	195	12	US-09-886-349A-8	Sequence 8, Appl
36	981	57.8	195	14	US-10-098-732A-8	Sequence 8, Appl
37	676	39.8	231	9	US-09-287-849-28	Sequence 28, Appl
38	676	39.8	231	14	US-10-359-460-28	Sequence 28, Appl
39	670	39.5	132	9	US-09-759-143-819	Sequence 819, App
40	670	39.5	132	9	US-09-759-143-848	Sequence 848, App
41	670	39.5	132	9	US-09-780-669-819	Sequence 819, App
42	670	39.5	132	9	US-09-780-669-848	Sequence 848, App
43	670	39.5	132	9	US-09-822-827-819	Sequence 819, App
44	670	39.5	132	9	US-09-822-827-848	Sequence 848, App
45	670	39.5	132	9	US-09-895-793-819	Sequence 819, App

ALIGNMENTS

RESULT 1

US-09-886-349A-4
; Sequence 4, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35 mature)
US-09-886-349A-4

Query Match 100.0%; Score 1698; DB 12; Length 330;

Best Local Similarity 100.0%; Pred. No. 1.1e-121;

Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MHHHHHAPPALSQRDFADFALPDP	SAVMAVQGVVNTKLGYNNAVGAGT	GIVID 60
DB	1	MHHHHHAPPALSQRDFADFALPDP	SAVMAVQGVVNTKLGYNNAVGAGT	GIVID 60
QY	61	PNGVVLTNNHVIAGATDINAFSGQ	TYGVVDVQVDTQDVAVLQLRAGGL	PSAAITGG 120
DB	61	PNGVVLTNNHVIAGATDINAFSGQ	TYGVVDVQVDTQDVAVLQLRAGGL	PSAAITGG 120

QY 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
DB 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDSGGPVVNGLGQVVMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDSGGPVVNGLGQVVMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 2

US-10-098-732A-4
; Sequence 4, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Ra35 mature
US-10-098-732A-4

Query Match 100.0%; Score 1698; DB 14; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.le-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGG 120
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGG 120
QY 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
DB 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDSGGPVVNGLGQVVMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDSGGPVVNGLGQVVMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 3

US-10-369-983-19
; Sequence 19, Application US/10369983
; Publication No. US2003023593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:wild-type
; OTHER INFORMATION: mature MTB32A (Ra35)
US-10-369-983-19

Query Match 100.0%; Score 1698; DB 15; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.le-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGG 120
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGG 120
QY 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
DB 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDSGGPVVNGLGQVVMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDSGGPVVNGLGQVVMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 4

US-09-886-349A-6
; Sequence 6, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Ra35FLMutSA
US-09-886-349A-6

Query Match
Best Local Similarity 99.8%; Score 1695; DB 12; Length 330;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAVGAGTGIVID 60
Db 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAVGAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120
Db 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120
QY 121 GVAAGEPVVAMGNSGGQGTFRAPVGRVVALGQTVQASDSLTAABETLNLGIQFDDAAIQP 180
Db 121 GVAAGEPVVAMGNSGGQGTFRAPVGRVVALGQTVQASDSLTAABETLNLGIQFDDAAIQP 180
QY 181 GDSGPPVNVGLGQVVGWNTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDSGPPVNVGLGQVVGWNTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 6
US-10-369-983-20
; Sequence 20, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB32AMutSA (Ra35 mutSA)
US-10-369-983-20

Query Match
Best Local Similarity 99.8%; Score 1695; DB 15; Length 330;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAVGAGTGIVID 60
Db 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAVGAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120
Db 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120
QY 121 GVAAGEPVVAMGNSGGQGTFRAPVGRVVALGQTVQASDSLTAABETLNLGIQFDDAAIQP 180
Db 121 GVAAGEPVVAMGNSGGQGTFRAPVGRVVALGQTVQASDSLTAABETLNLGIQFDDAAIQP 180
QY 181 GDSGPPVNVGLGQVVGWNTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDSGPPVNVGLGQVVGWNTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Ra35FLMutSA
US-10-098-732A-6

Query Match
Best Local Similarity 99.8%; Score 1695; DB 14; Length 330;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAVGAGTGIVID 60
Db 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAVGAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120
Db 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120

RESULT 5
US-10-098-732A-6
; Sequence 6, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Ra35FLMutSA
US-10-098-732A-6
```

RESULT 7
 US-10-369-983-2
 ; Sequence 2, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 723
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: mutated
 ; OTHER INFORMATION: MTB32-MTB39F fusion protein (MTB32MutSA)
 US-10-369-983-2

Query Match 99.8%; Score 1695; DB 15; Length 723;
 Best Local Similarity 99.7%; Pred. No. 4.6e-121;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGAGTGIVID 60
 Db 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGAGTGIVID 60
 QY 61 PNGVVLNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLGAGLPSAAIGG 120
 Db 61 PNGVVLNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLGAGLPSAAIGG 120
 QY 121 GVAVGEPVVMGNSGGGCTPRVGRVVALGQTVQASDSLTCABETLNGLIQFDDAATOP 180
 Db 121 GVAVGEPVVMGNSGGGCTPRVGRVVALGQTVQASDSLTCABETLNGLIQFDDAATOP 180
 QY 181 GDSGPPVNGLGQVVGNNMTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 Db 181 GDAGGPPVNGLGQVVGNNMTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
 Db 241 PTAFLGLGVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 8
 US-10-369-983-4
 ; Sequence 4, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 1010
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: MTB-102F fusion
 ; OTHER INFORMATION: protein
 US-10-369-983-4

Query Match 99.8%; Score 1695; DB 15; Length 1010;
 Best Local Similarity 99.7%; Pred. No. 7e-121;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGAGTGIVID 60
 Db 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGAGTGIVID 60
 QY 61 PNGVVLNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLGAGLPSAAIGG 120
 Db 61 PNGVVLNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLGAGLPSAAIGG 120
 QY 121 GVAVGEPVVMGNSGGGCTPRVGRVVALGQTVQASDSLTCABETLNGLIQFDDAATOP 180
 Db 121 GVAVGEPVVMGNSGGGCTPRVGRVVALGQTVQASDSLTCABETLNGLIQFDDAATOP 180
 QY 181 GDSGPPVNGLGQVVGNNMTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 Db 181 GDAGGPPVNGLGQVVGNNMTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
 Db 241 PTAFLGLGVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 9
 US-09-712-363-161
 ; Sequence 161, Application US/09712363
 ; Patent No. US20020164588A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eisenberg, David
 ; APPLICANT: Rotstein, Sergio H.
 ; APPLICANT: Marcotte, Edward M.
 ; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
 ; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
 ; FILE REFERENCE: 07419-032001
 ; CURRENT APPLICATION NUMBER: US/09/712,363
 ; CURRENT FILING DATE: 2000-11-13
 ; PRIOR APPLICATION NUMBER: PCT/US00/02246
 ; PRIOR FILING DATE: 2000-01-28
 ; PRIOR APPLICATION NUMBER: 60/179,531
 ; PRIOR FILING DATE: 2000-02-01
 ; PRIOR APPLICATION NUMBER: 60/117,844
 ; PRIOR FILING DATE: 1999-01-29
 ; PRIOR APPLICATION NUMBER: 60/118,206,
 ; PRIOR FILING DATE: 1999-02-01
 ; PRIOR APPLICATION NUMBER: 60/126,593
 ; PRIOR FILING DATE: 1999-03-26
 ; PRIOR APPLICATION NUMBER: 60/134,093
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/134,092
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/165,124
 ; PRIOR FILING DATE: 1999-11-12
 ; PRIOR APPLICATION NUMBER: 60/165,086
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 292
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 161

LENGTH: 355
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-161

Query Match
Best Local Similarity 96.2%; Score 1645; DB 9; Length 355;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVAGTGIVIDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVAGTGIVIDPNGVLT 92
QY 68 NNHVIAGATDINAFSVGSGQTYGVYVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVGSGQTYGVYVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNGLGQVGVNMTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGL 247
DB 213 VNGLGQVGVNMTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGL 272
QY 248 GVVDNNGNGARVQVVGSAASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLLAEGPPA 330
DB 333 VTWQTKSGGTRTGNVTLLAEGPPA 355

RESULT 10
US-09-886-349A-2
Sequence 2, Application US/09886349A
Publication No. US20040086523A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Reed, Steven
APPLICANT: Alderson, Mark
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-0090700S
CURRENT APPLICATION NUMBER: US/09/886,349A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: US 09/597,796
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: US 60/265,737
PRIOR FILING DATE: 2001-02-01
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 355
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: MTB32A (Ra35FL)
US-09-886-349A-2

Query Match
Best Local Similarity 99.4%; Score 1634; DB 12; Length 355;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVAGTGIVIDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVAGTGIVIDPNGVLT 92
QY 68 NNHVIAGATDINAFSVGSGQTYGVYVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVGSGQTYGVYVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 152

QY 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNGLGQVGVNMTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGL 247
DB 213 VNGLGQVGVNMTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGL 272
QY 248 GVVDNNGNGARVQVVGSAASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLLAEGPPA 330
DB 333 VTWQTKSGGTRTGNVTLLAEGPPA 355

RESULT 11
US-10-193-002-80
Sequence 80, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campese-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-10-193-002-80

Query Match
Best Local Similarity 96.2%; Score 1634; DB 14; Length 355;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAGAGTGIVDPNGVVL 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAGAGTGIVDPNGVVL 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNGLGQVVGMNTAASDNFQLSGGGQGFPAIPGQAWAIGQIRSGGSPVTHIGPTAFGL 247
DB 213 VNGLGQVVGMNTAASDNFQLSGGGQGFPAIPGQAWAIGQIRSGGSPVTHIGPTAFGL 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 12

US-10-084-843-79
; Sequence 79, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campes-Neto, Antonio
; Houghton, Raymond
; Wedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-084-843-79

Query Match 96.2%; Score 1634; DB 14; Length 355;
Best Local Similarity 99.4%; Pred. No. 8.9e-117;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAGAGTGIVDPNGVVL 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAGAGTGIVDPNGVVL 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNGLGQVVGMNTAASDNFQLSGGGQGFPAIPGQAWAIGQIRSGGSPVTHIGPTAFGL 247
DB 213 VNGLGQVVGMNTAASDNFQLSGGGQGFPAIPGQAWAIGQIRSGGSPVTHIGPTAFGL 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 13

US-10-098-732A-2
; Sequence 2, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 355
; TYPE: PPT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35FL)
US-10-098-732A-2

Query Match 96.2%; Score 1634; DB 14; Length 355;
Best Local Similarity 99.4%; Pred. No. 8.9e-117;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAGAGTGIVDPNGVVL 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNNINIKLGYNNVAGAGTGIVDPNGVVL 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 187

Db 153 VVAMGNSGGGTPRAVPGRVVVALGQTVAQSDSLTGABETLNGLIQFDAAIQPGDSGGPV 212
Qy 188 VNLGQVVGWNTAAADNFQLSQGGQGFALPQGMAMIAQIBSGSGSPTVHIGPTAFGL 247
Db 213 VNLGQVVGWNTAAADNFQLSQGGQGFALPQGMAMIAQIBSGSGSPTVHIGPTAFGL 272
Qy 248 GVVDNNGNGARVQVVGSAASLIGTGDVITAVDVGAPINSATAMADALNGHHFGDVIS 307
Db 273 GVVDNNGNGARVQVVGSAASLIGTGDVITAVDVGAPINSATAMADALNGHHFGDVIS 332
Qy 308 VVWQTKSGGTRGNTVLAEGPPA 330
Db 333 VVWQTKSGGTRGNTVLAEGPPA 355

RESULT 14
US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 59.1%; Score 1003.5; DB 15; Length 1016;
Best Local Similarity 85.9%; Pred. No. 4.6e-68;
Matches 207; Conservative 6; Mismatches 13; Indels 15; Gaps 4;
Qy 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGTGVIDPNGVLT 67
Db 535 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGTGVIDPNGVLT 594
Qy 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVQLRGAGGLPSAAIGGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVQLRGAGGLPSAAIGGGVAVGEP 654
Qy 128 VVAMGNSGGGTPRAVPGRVVVALGQTVAQSDSLTGABETLNGLIQFDAAIQPGDSGGPV 187
Db 655 VVAMGNSGGGTPRAVPGRVVVALGQTVAQSDSLTGABETLNGLIQFDAAIQPGDSGGPV 714
Qy 188 VNLGQVVGWNTAAADNFQLSQGGQGFALPQGMAMIAQIBSGSGSPTVHIGPTAFGL 238
Db 715 VNLGQVVGWNTAAASGTF--SRFG-----LPEVYLVQVPSWGRDVKVQFSGGNNSPAVY 768
Qy 239 I 239
Db 769 L 769

RESULT 15
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-16

Query Match 58.6%; Score 995; DB 15; Length 1154;
Best Local Similarity 73.2%; Pred. No. 2.4e-67;
Matches 213; Conservative 13; Mismatches 37; Indels 28; Gaps 4;
Qy 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGTGVIDPNGVLT 67
Db 535 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGTGVIDPNGVLT 594
Qy 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVQLRGAGGLPSAAIGGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVQLRGAGGLPSAAIGGGVAVGEP 654
Qy 128 VVAMGNSGGGTPRAVPGRVVVALGQTVAQSDSLTGABETLNGLIQFDAAIQPGDSGGPV 187
Db 655 VVAMGNSGGGTPRAVPGRVVVALGQTVAQSDSLTGABETLNGLIQFDAAIQPGDSGGPV 714
Qy 188 VNLGQVVGWNTAAADNFQLSQGGQGFALPQGMAMIAQIBSGSGSPTVHIGPTAFGL 247
Db 715 VNLGQVVGWNTAAASGTFMDF-----GLLPP-----EVNSSRMYSGPESMLAAAAWDGV 765
Qy 248 GVVDNNGNGARVQVVGSAASLIGTGDVITAVDVGAPI--NSATAMADA 296
Db 766 -----AAELTSAVSYGSVVSTLIVEPMWGPAAAAAAA 799

Search completed: June 22, 2004, 18:07:52
Job time : 31.5644 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 9.233 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSDQRFADP.....QTKSGGTRGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: **
1: pir1: **
2: pir2: **
3: pir3: **
4: pir4: **

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1645	96.9	355	2 F70983	probable serine pr
2	1231.5	72.5	361	2 S47170	hypothetical prote
3	1198	70.6	354	2 A87242	probable secreted
4	459.5	27.1	464	2 C70821	probable serine pr
5	450	26.5	382	2 H86930	probable secreted
6	450	26.5	452	2 T45448	probable serine pr
7	374	22.0	394	2 S74643	proteinase hhoA (E
8	368	21.7	407	2 AG2150	serine proteinase
9	333	20.8	433	2 H97199	htrA-like serine p
10	332	20.7	452	2 S77538	serine proteinase
11	338	19.9	514	2 A82581	periplasmic protei
12	335.5	19.8	362	2 T35287	probable secreted
13	333	19.6	416	2 AB2057	serine proteinase
14	332	19.6	398	2 B71284	probable periplasm
15	332	19.6	441	2 E75357	probable periplasm
16	329.5	19.4	429	2 AD1894	serine proteinase
17	319.5	18.8	474	2 F83550	serine proteinase
18	319	18.8	355	1 JC6052	trypsin-like prote
19	319	18.8	355	2 D91142	trypsin-like prote
20	318	18.7	355	2 G85987	trypsin-like prote
21	317.5	18.7	401	2 AD2451	serine proteinase
22	315	18.6	408	2 H86891	exported serine pr
23	313.5	18.5	352	2 E82307	trypsin-like prote
24	312.5	18.4	455	2 AB0909	serine protease (E
25	312	18.4	416	2 S75445	proteinase hhoB (E
26	310.5	18.3	348	2 H96956	serine protease Do
27	310.5	18.3	455	2 C91142	serine endoprotein
28	309	18.2	455	2 F85987	serine endoprotein
29	308.5	18.2	513	2 AD3418	proteinase DO (EC

30 307.5 18.1 513 2 I40060 serine proteinase
31 306.5 18.1 474 1 I40059 htrA-like protein
32 306.5 18.1 474 2 A13349 proteinase DO (EC
33 305.5 18.0 455 2 JC6051 trypsin-like prote
34 304 17.9 328 1 D69109 serine proteinase
35 302.5 17.8 356 2 AC0909 trypsin-like prote
36 301 17.7 459 2 F72359 periplasmic serine
37 299 17.6 457 2 AG0433 proteinase (EC 3.4
38 298 17.6 499 2 H81914 probable periplasm
39 296 17.4 475 1 S15337 heat shock protein
40 294.5 17.3 363 2 C87336 serine proteinase
41 294.5 17.3 449 2 A69643 serine proteinase
42 294.5 17.3 474 2 S45229 proteinase DO (EC
43 294.5 17.3 474 2 B85500 proteinase DO (EC
44 294.5 17.3 474 2 E90649 proteinase DO (EC
45 294 17.3 476 2 H71936 proteinase DO - He

ALIGNMENTS

RESULT 1

F70983
probable serine proteinase: pepA - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70983

R:Coile, S.I.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70983

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-355 <COL>

A:Cross-references: GB:Z96071; GB:AL123456; NID:g3242254; PIDN:CAB09453.1; PID:g2181967

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: pepA

C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypsin

Query Match 96.9%; Score 1645; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 9.4e-90;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSDQRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVIDPENGVVLT 67

Db 33 APPALSDQRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVIDPENGVVLT 92

QY 68 NNHVIAGATDINAFSVSGSGQTVGVVGYDRDTQDVAVILQLRGAGGLPSAAIGGGVAVGCEP 127

Db 93 NNHVIAGATDINAFSVSGSGQTVGVVGYDRDTQDVAVILQLRGAGGLPSAAIGGGVAVGCEP 152

QY 128 VVAMGNSGGGCTPRAVGRVVALGQTVQASDSITGAEETLNGLIQFDAAIQPDGSGGPV 187

Db 153 VVAMGNSGGGCTPRAVGRVVALGQTVQASDSITGAEETLNGLIQFDAAIQPDGSGGPV 212

QY 188 VNLGQVQVGMNTAASDNFQLSQGGGFAIPGQAWATAGQIRSGGSGTTHIGTAFGL 247

Db 213 VNLGQVQVGMNTAASDNFQLSQGGGFAIPGQAWATAGQIRSGGSGTTHIGTAFGL 272

QY 248 GVDDNNGNGARVQVWGSAPAAISLGISTGDIVITAVDGAIPNSATAMADALNGHPGDVIS 307

Db 273 GVDDNNGNGARVQVWGSAPAAISLGISTGDIVITAVDGAIPNSATAMADALNGHPGDVIS 332

QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330

Db 333 VTWQTKSGGTRTGNVTLAEGPPA 355

Db 436 APGATVALTFQDPGSGSRTVQVTLGK 461

RESULT 5

H86930

Probable secreted serine proteinase [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002

C;Accession: H86930

R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

A;Title: Massive gene decay in the leprosy bacillus.

A;Reference number: A86909; MUID:21128732; PMID:11234002

A;Accession: H86930

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-382 <STO>

A;Cross-references: GB:AL450380; NID:gl3092536; PIDN:CAC29684.1; GSPDB:GNC0147

C;Genetics:

A;Gene: M0176

C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 26.5%; Score 450; DB 2; Length 382;

Best Local Similarity 37.5%; Pred. No. 1.4e-19;

Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALPLDPSAMVAQVG-----PQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 73

Db 59 ASVPAANM-PSGSVEQAVKVPVSMLETLGRQSE--EGSGVILSADGLILTNHVA 115

QY 74 GA-----TDINAFSGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIG 119

Db 116 VAAKPGGGGGLSPKTTVTFF---DGTASFVVGADPTSDIAVVRVQSISGLTPTIMG 172

QY 120 GG--VAVGEPPVAMNSGGGGTTPRAVGRVVALGTQVQASDLSLTGAEETLNGLIQFDAA 177

Db 173 SSADLRVQGPVAVGSPGLAGT--VTSGIVSALNRPVSTTGE--SGNQTVDLAIQTDA 229

QY 178 IQPGDSGGPVVNGLGQVGMNTAA-----SDNFQLSQGGQFAIPGQAMAIAGQIRSG 231

Db 230 INFGSGGALVNMGGQLVGVNSAIATLGADSGAQSGSIGLGFPAIPVDQAKRIADELIST 289

QY 232 GGSPTVHIGPTAFGLGVDDNNGN--GARVQRVVGSAASLSLTGCVITAVDVGAPINSA 290

Db 290 G--KATH---ASLGQVATDKGTGPAKVMVDVAGGAANAAPKGVVLTKVDRLLISSA 343

QY 291 TAMADALNGHHPGDVSVTWQTSKGGTRTGNVTLAE 326

Db 344 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 379

RESULT 6

T45448

Probable serine proteinase (EC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Dec-2002

C;Accession: T45448

R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, February 1998

A;Reference number: Z22967

A;Accession: T45448

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-452 <JAM>

A;Cross-references: EMBL:AL035500; PIDN:CAB36690.1

A;Experimental source: cosmid L373

C;Genetics:

A;Gene: MLCB373.28

C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

C;Keywords: hydrolase; serine proteinase

Query Match 26.5%; Score 450; DB 2; Length 382;

Best Local Similarity 37.5%; Pred. No. 1.4e-19;

Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALPLDPSAMVAQVG-----PQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 73

Db 59 ASVPAANM-PSGSVEQAVKVPVSMLETLGRQSE--EGSGVILSADGLILTNHVA 115

QY 74 GA-----TDINAFSGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIG 119

Db 116 VAAKPGGGGGLSPKTTVTFF---DGTASFVVGADPTSDIAVVRVQSISGLTPTIMG 172

QY 120 GG--VAVGEPPVAMNSGGGGTTPRAVGRVVALGTQVQASDLSLTGAEETLNGLIQFDAA 177

Db 173 SSADLRVQGPVAVGSPGLAGT--VTSGIVSALNRPVSTTGE--SGNQTVDLAIQTDA 229

QY 178 IQPGDSGGPVVNGLGQVGMNTAA-----SDNFQLSQGGQFAIPGQAMAIAGQIRSG 231

Db 230 INFGSGGALVNMGGQLVGVNSAIATLGADSGAQSGSIGLGFPAIPVDQAKRIADELIST 289

QY 232 GGSPTVHIGPTAFGLGVDDNNGN--GARVQRVVGSAASLSLTGCVITAVDVGAPINSA 290

Db 290 G--KATH---ASLGQVATDKGTGPAKVMVDVAGGAANAAPKGVVLTKVDRLLISSA 343

QY 291 TAMADALNGHHPGDVSVTWQTSKGGTRTGNVTLAE 326

Db 344 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 379

F.182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 26.5%; Score 450; DB 2; Length 452;

Best Local Similarity 37.5%; Pred. No. 1.7e-19;

Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALPLDPSAMVAQVG-----PQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 73

Db 129 ASVPAANM-PSGSVEQAVKVPVSMLETLGRQSE--EGSGVILSADGLILTNHVA 185

QY 74 GA-----TDINAFSGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIG 119

Db 186 VAAKPGGGGGLSPKTTVTFF---DGTASFVVGADPTSDIAVVRVQSISGLTPTIMG 242

QY 120 GG--VAVGEPPVAMNSGGGGTTPRAVGRVVALGTQVQASDLSLTGAEETLNGLIQFDAA 177

Db 243 SSADLRVQGPVAVGSPGLAGT--VTSGIVSALNRPVSTTGE--SGNQTVDLAIQTDA 299

QY 178 IQPGDSGGPVVNGLGQVGMNTAA-----SDNFQLSQGGQFAIPGQAMAIAGQIRSG 231

Db 300 INFGSGGALVNMGGQLVGVNSAIATLGADSGAQSGSIGLGFPAIPVDQAKRIADELIST 359

QY 232 GGSPTVHIGPTAFGLGVDDNNGN--GARVQRVVGSAASLSLTGCVITAVDVGAPINSA 290

Db 360 G--KATH---ASLGQVATDKGTGPAKVMVDVAGGAANAAPKGVVLTKVDRLLISSA 413

QY 291 TAMADALNGHHPGDVSVTWQTSKGGTRTGNVTLAE 326

Db 414 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 449

RESULT 7

S74643

Proteinase rhoA (EC 3.4.-) - Synecocystis sp. (strain PCC 6803)

N;Alternate names: protein sll1679

C;Species: Synecocystis sp.

A;Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, D.N.A. Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis s.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S74643

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-394 <KAN>

A;Cross-references: EMBL:D90900; GB:AB001339; NID:gl651768; PIDN:BAAL6795.1; PID:gl651861

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Gene: rhoA

C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

C;Keywords: hydrolase; proteinase

Query Match 22.0%; Score 374; DB 2; Length 394;

Best Local Similarity 31.6%; Pred. No. 4.3e-11;

Matches 118; Conservative 49; Mismatches 111; Indels 96; Gaps 15;

QY 9 PPA-----LSQDRFADFPPALPLDPSAMVAQVGPVVNNTK-----44

Db 39 PPAPVITAAQASVPLTSESFV-----AAVSRSGEAVVVIDTETVVTTRTDILDDP 89

QY 45 -----LGNNNAV-----GAGTGIVDPNGVLTNNHVIAGATDINAFVSGSQTYGV 91

Db 90 PFQEPFGRSFPVPRERRIAGGSGFIIDNSGIILTAHVVDGASKV-VVTLRDGRTFDG 148

QY 92 DVVGYDRTQDVAVLQLRGAG-GLPSAAIG--GGVAVGPPVAMNSGGGGTTPRAVGRV 148

Db 149 QVRGDEVTDLAVAVKIEPQGGALPVAPLGTSNQLQVGDAIAGNVPGLDNT-----200

QY 149 VALG--QTVAQSDSLTGAEETLNGLIQFDAAIQPDGSGPPVNGLGQVVGMTAASDNFQ 206

QY 293 MADALNGHPGDVISV 308
:
: : : :
Db 400 IDSISKNSGDTQV 415

RESULT 10
S77538
serine proteinase (EC 3.4.21.-) htrA - Synchocystis sp. (strain PCC 6803)
N:Alternate names: protein slr1204
C:Species: Synchocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 htext_change 09-Dec-2002
C:Accession: S77538
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S77538
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-452 <RAN>
A:Cross-references: EMBL:D90905; GB:AB001339; NID:G1652360; PIDN:BAAL17385.1; PID:G1652463;
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: htrA
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypsin
C:Keywords: hydrolase; serine proteinase

C;Keywords: hydrolase; serine proteinase

Query Match	20.7%;	Score 352;	DB 2;	Length 452;
Best Local Similarity	30.2%;	Pred. No. 9.8e-14;		
Matches 110;	Conservative 59;	Mismatches 113;	Indels 82;	Gaps 15;

QY 24 PLDPAMVAQV---GPOVNTNKLGVNNAV-----G 52
 Db 112 PREPSNFVVDVSTGPAVVRINAQTKVSKQVPOAFNDPFLQRFQSQPMPNERNVQRG 171
 QY 53 ACTGVVIDPVGVLVNNHVIAGATDINAFSGSGGTQYGVVGYDRTODVAVLQRLGAGG 112
 Db 172 TSGGFIIVNDGKIFITNAHVVDGAEVTV-TLKDGRSFCRVMGSPDSDVAVVKE-AGD 229
 QY 113 LPSAAIGGG--VAVGEPVVMNGSGGGGTTPRAVGRVVALGQTVQASDLSLTGAEETLNG 170
 Db 230 LPTVALGDSHDLQVGWEAIAIGNPLGLDNT--VTTGILSATGR--RSAD--IGVDPKRE 283
 QY 171 LTFQDAAIOPSGSGPVNGLGVVGMNTAASDNFQLSGOGGFAIPGQAMAIAGQIRS 230
 Db 284 FIQTDAIINPNSGGPGLNADGGVGMNTAIONAQ-----GIGFALPINKAQEIIAQJIA 339
 QY 231 GGGSPFVTHIGPTAFLGLGVVDNN-----GNGARVQRVVGSAAPASLGI 273
 Db 340 TG--KVEH---AYLGIQVMTVTEPQLSQIRQETGMNI PVDXGVVIMQVMPNSPAAIAKL 393
 QY 274 STGDVITAVDGPINATAMADALNGHPGDVIVTWQTKSGGTRTG---NVTLAGG--- 327
 Db 394 EQGDVLQSLQGGPVNAEQVQLVGLAVGDEVEL-----GILRNGQOONLTVTIGALP 447
 QY 328 --PP 329
 Db 448 SAPP 451
 RESULT 11
 A:Gene: Xf2241 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A82581
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A;Reference number: A82515; PMID:20365717; PMID:10910347
 A;Note: for a complete list of authors see reference number A59328 below
 A;Accession: A82581
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-514 <SIM>
 A;Cross-references: GB:AE003849; NID:g9107394; PIDN:AAF85040.1; GSPDB:GN001
 A;Experimental source: Strain 9a5c
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 ; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A;Reference number: A59328
 A;Contents: annotation
 C;Genetics:
 A;Gene: Xf2241
 C;Superfamily: Helicobacter serine proteinase
 Query Match 19.9%; Score 338; DB 2; Length 514;
 Best Local Similarity 32.6%; Pred. No. 7.5e-13;
 Matches 109; Conservative 48; Mismatches 107; Indels 70; Gaps 13;
 QY 23 LPLDPSAMVAQVGPQVNNINT-----KLGVNNAV----- 51
 Db 54 LP-DFTQLVDQVGPQVNNIETVTRKKVGRGRIPLNDIPEFFRFRFGPDMQNPQRG 112
 QY 52 -----GAGTGIVIDPVGVLVNNHVIAGATDINAFSGSGGTQYGVVGYDRTQ 100

Db 113 QODDEGGIAGMGSGFIISKDGYILTNHVIITGASEV-TIKLTDREFRAKIIGSEQY 171
 QY 101 DVAVLQRLGAGGIPSAAIAGGVAV--GEPVVMNGSGGGGTTPRAVGRVVALGQTVQAS 158
 Db 172 DVALLKI-DARNLFTVRIGDSSSLKQGVVAIGSPFGLDHSVTA--GIVSALGRS----- 224
 QY 159 DLSLTGAEETLNGLIQFDAAIOPGDSGGPVNGLGVVGMNTAASDNFQLSQG--GQGFAL 216
 Db 225 ---TSDQRYVFPFQTDVPINQNGSGGPLNTRGEVIGIN---SQIFASGGYMGISFAI 278
 QY 217 PIGQAMAIAGQIRSGG---GSPFVTHIGPTAFL---GLGVVDNNGCARVQRVVGSAAP 269
 Db 279 PINLAINAAEQIRKTKGVQSRMLGVEIGPIDALKAQGLGLPDS--RGALVNNIIPHPSPA 336
 QY 270 SLGISTGDVITAVDGPINATAMADALNGHPG 303
 Db 337 KAGIEVGDVIRSVNGKVIVSSFDLPPLIGMPPG 370
 RESULT 12
 T35287
 probable secreted proteinase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Dec-2002
 C:Accession: T35287
 R;Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A;Reference number: Z21574
 A;Accession: T35287
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-362 <SE>
 A;Cross-references: EMBL:AL096872; PIDN:CAB51255.1; GSPDB:GN00070; SCODEB:SC5F7.30
 A;Experimental source: strain A3(2)
 C;Genetics:
 A;Gene: SCODEB:SC5F7.30
 C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps
 Query Match 19.8%; Score 335.5; DB 2; Length 362;
 Best Local Similarity 33.9%; Pred. No. 7.2e-13;
 Matches 113; Conservative 45; Mismatches 138; Indels 37; Gaps 12;
 QY 8 APPALSDQRFADFPALPDPDPMVAQVGPQVNNIKNLGVNNAVAGAGTIVIDPNGVLT 67
 Db 52 APRAASE-----LEADYERIVKDVLPVSVVQIA-----GDSLGSVVYDDKGHVVT 97
 QY 68 NNHVIAGATDINAFSGSGGTQYGVVGYDRTQDVAVLQL-RGAGGLPSA--AIGGG 121
 Db 98 NHHVVG---DAKSFVTTARTEGALTAKLVSSYPQDLAVIKLDKVPESMRAARFADSAK 154
 QY 122 VAVGSPVVMNGSGGGGTTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIOPG 181
 Db 155 VEVGQIVLAMGSPGLSSS--VTQIVSATGRTVTEGSSGGGTGATIANMVQTSAAINPG 212
 QY 182 DSGGPVNVGLGVVGMNTAASDNFQLSQG---GQGFALPIGQAMAIAGQIRSGG---GSP 235
 Db 213 NSGGALVNDGGQVIGIFLTAIDPGLGSAAPGIGFALPASMTVVAQIVRDGKVTDSG 272
 QY 236 TVHIGTAPFLGLGVVDNNGN--GARVQRVVGSAAPASLGISTDVITAVDGPINATAM 293
 Db 273 RAALGITA---RTVDDSVRPAGAAVVEVSDGGAADAGLRPGDVLVGLGIDTITITSL 329
 QY 294 ADALNGHPGDVIVTWQTKSGGTETGNVTLAE 326
 Db 330 SEALASMRPGRDRTKVY-TRDGKHTAEVTLGE 361
 RESULT 13
 AB2057
 serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

Db 371 LPSAQTSISTDGLITAVNGQPLEDAGSLQEAVALATGEGQPLRLTVR-RGGKTREVEV 429

QY 323 TL 324

Db 430 TL 431

Search completed: June 22, 2004, 17:24:48
Job time : 10.233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 5.49204 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSDQRFADF.....QTKSGGTRTGNVTLAEGPPA 330

Scoring table: BLOSOM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	326.5	19.2	458	1	YVTA_BACSU	Q9311 bacillus su
2	326	19.2	413	1	HTRA_LACHE	Q924n7 lactobacill
3	319	18.8	355	1	DEGS_ECOLI	P31137 escherichia
4	315	18.6	408	1	HTRA_LACLA	Q91a06 lactococcus
5	308.5	18.2	513	1	DEGP_BRUME	C8Y932 brucella me
6	307.5	18.1	513	1	DEGP_BRUSU	Q44597 brucella su
7	305.5	18.0	435	1	DEGO_ECOLI	P39099 escherichia
8	296	17.4	475	1	DEGP_SALTY	P26382 salmonella
9	296	17.4	504	1	DEGP_RHIME	Q52894 rhizobium m
10	294.5	17.3	449	1	HTRA_BACSU	Q34358 bacillus su
11	294.5	17.3	474	1	DEGP_ECOLI	P09376 escherichia
12	292	17.2	466	1	HTOA_HABIN	P45129 haemophilus
13	288.5	17.0	503	1	DEGP_BABHE	P54925 bartonella
14	288	17.0	497	1	DEGP_CHLTR	P18594 chlamydia t
15	287	16.9	478	1	DEGP_BUCAI	P57322 buchnera ap
16	283.5	16.7	497	1	DEGP_CHLMU	Q9p197 chlamydia m
17	282	16.6	489	1	DEGP_CHLPN	Q92609 arabidopsis
18	281	16.5	437	1	DEGL_ARATH	Q921a1 rickettsia
19	280	16.5	508	1	DEGP_RICCN	P39668 bacillus su
20	274.5	16.2	400	1	YVXA_BACSU	Q91u10 arabidopsis
21	272.5	16.0	448	1	DEG8_ARATH	Q85291 buchnera ap
22	270.5	15.9	478	1	DEGP_BUCAP	O05942 rickettsia
23	254	15.0	513	1	DEGP_RICPR	O43464 homo sapien
24	248.5	14.6	458	1	HRA2_HUMAN	P44947 haemophilus
25	242	14.3	340	1	DEGS_HABIN	Q92743 homo sapien
26	235.5	13.9	480	1	HEAL_HUMAN	P83110 homo sapien
27	234.5	13.8	453	1	HRA3_HUMAN	Q9j115 mus musculu
28	228.5	13.5	458	1	HRA2_MOUSE	P83105 homo sapien
29	225	13.3	476	1	HRA4_HUMAN	Q9d236 mus musculu
30	221.5	13.0	460	1	HRA3_MOUSE	Q9r118 mus musculu
31	221	13.0	480	1	HRA1_MOUSE	Q9sel7 arabidopsis
32	178.5	10.5	321	1	SPPA_ARATH	P55377 rhizobium s
33	175.5	10.3	630	1	Y4BU_RHISN	

ALIGNMENTS

RESULT 1

YVTA_BACSU STANDARD; PRT; 458 AA.

34	151	8.9	747	1	ELS_BOVIN
35	150.5	8.9	710	1	FIBI_ADEG1
36	150	8.8	457	1	PRTC_STRGR
37	149.5	8.8	515	1	YI40_MYCTU
38	146.5	8.6	864	1	ELS_RAT
39	145.5	8.6	957	1	Y278_MYCTU
40	143	8.4	801	1	Y747_MYCTU
41	141	8.3	2021	1	OMPA_RICCN
42	140	8.2	1150	1	APMU_FIG
43	138	8.1	914	1	WA22_MYCTU
44	137	8.1	778	1	YC3*_MYCTU
45	136	8.0	2249	1	OMPA_RICRI

P04985	bos taurus
Q64761	avian adeno
P52320	streptomyce
Q50594	mycobacteri
Q99372	rattus norv
P56877	mycobacteri
Q53810	mycobacteri
Q52657	rickettsia
P12021	sus scrofa
Q06794	mycobacteri
P71933	mycobacteri
P15921	rickettsia

AC Q9R91I; Q35021; Q35039;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease yvta (EC 3.4.21.-).
GN YVTA OR YVTB OR BSU33000.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykda, encoding a Bacillus subtilis homologue of HtrA,
RT is heat shock inducible and negatively autoregulated";
RL J. Bacteriol. 182:1592-1599(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98015415; PubMed=9353931;
RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and citG (289
degrees) in Bacillus subtilis";
RL Microbiology 143:3305-3308(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bortner M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Ertz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.F., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi H., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pehl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,


```
Matches 90; Conservative 50; Mismatches 121; Indels 24; Gaps 8;
QY 42 NTKLGYNAVAGTGIIV--DPNGVLTNNHVIAGATDINAFSGVGQTGVGVVGYDR 98
Db 119 SSXNGKLEYSSEGGVYVYKSGKGIYVNNHVISGSDAVQVL-LANGKTVNAKVVGKDS 177
QY 99 TQDAVLQIRAGGLPSAIGGG--VAVGEPVVMNGSGGGGTTPRAVPGRVVVLGQTVQ 156
Db 178 TTDLAVLSDAKYVTTQAGFGSKHLEAGQTVIAGVSPILGSEYASTVTQGIISAPARTIS 237
QY 157 ASDSLTGAETLNLGLQFADAAITOPGSGGVPVNGVGQVVGMMT--AASDNFQSGGGQF 214
Db 238 TS---SGNQTT---VLQTDALNPNNGSGALVNSAGQVIGNSMKLAQSSDTSVEGNAP 291
QY 215 AIPIGQAMAIAAGIRSGG--GSFTVHIGTAFLG-----LGVVDNNGNGARVQRVVG 264
Db 292 AIPNSNEWTVNELVKKGKITRPLQGVGVIALQGIPEGVRSRLKIKSNLKNGIYIAFVSR 351
QY 265 SAPAASLGSTGDVITAVGAPINSATAMADALNGHHPGDVISVT 309
Db 352 NGSAAAGIKSGDVITKVDKGVEDVASLHSILYSHKVGDTNVNT 396

RESULT 3
DEGS ECOLI STANDARD; PRT; 355 AA.
ID DEGS_ECOLI
AC F31137;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease degs precursor (EC 3.4.21.-).
GN DEGS OR HHOB OR HTHF OR B3235 OR Z4594 OR EC94108.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Bass S., Gu Q., Goddard A.;
RL Submitted (0CT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;
RA Waller P.R., Sauer R.T.;
RT "Characterization of degQ and degS, Escherichia coli genes encoding
RT homologs of the DegP protease.";
RL J. Bacteriol. 178:1146-1153(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:12453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba I., Hattori M., Shinagawa H., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
RX MEDLINE=88105815; PubMed=3322223;
RA Vogel R.F., Entian K.-D., Mecke D.;
RT "Cloning and sequence of the mdh structural gene of Escherichia coli
RT coding for malate dehydrogenase.";
RL Arch. Microbiol. 149:36-42(1987).
RN [7]
RP IDENTIFICATION.
RA Bazan J.F., Fletterick R.J.;
RT "Structural and catalytic models of trypsin-like viral proteases.";
RL Semin. Virol. 1:311-322(1990).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2c.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; U15661; AAC43993.1; -.
CC EMBL; U32495; AAC4006.1; -.
CC EMBL; U18997; AAA58037.1; -.
CC EMBL; AE000402; AAC76267.1; -.
CC EMBL; AE005551; AAG58363.1; -.
CC EMBL; AP002564; BAB37531.1; -.
CC EMBL; M24777; -. NOT_ANNOTATED_CDS.
CC PIR; D91142; D91142.
CC PIR; JC6052; JC6052.
CC MEROPS; S01.275; -.
CC EcoGene; EG11652; hhoB.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ_Ser_trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1c.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS0106; PDZ; 1.
CC SIGNAL 1 28
KW Hydrolase; Serine protease; Periplasmic; Signal; Complete proteome.
FT SIGNAL 29 355
FT CHAIN 29 355
FT DOMAIN 281 326
FT ACT_SITE 96 96
FT ACT_SITE 126 126
FT ACT_SITE 201 201
FT CONFLICT 253 253
FT CONFLICT 307 307
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;
Query Match 18.8%; Score 319; DB 1; Length 355;
Best Local Similarity 31.2%; Pred. No. 1.9e-11;
Matches 109; Conservative 43; Mismatches 137; Indels 50; Gaps 13;
QY 12 LSQDFADFPALPDPSAMVAGVQVYVNTKGLYNNVAGV-----GTGVID 60
Db 32 LSTPQDSTDETPASVLAVERAAPVNV-----YNRGLNTNSHNLQIRLTGSGVMD 86
```

QY 61 PNGVLTNNHVIAGATDINAFSGGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIG 120
Db 87 QRGVITIKHVIINDAQI-IVALQGRVFEALLVGSLSLTLAVLKLINATGGGLTIPINA 145
QY 121 GVA--VVEPVVAMNSGGQGTTPRAVGRVVALGQTV-QASDSLTAABETLN-----GLI 172
Db 146 RRVPHIGDVLAINP-----YNLQGITQIGIISATG-RIGLNTGRQNFL 190
QY 173 QFDAIIPQSGGVPVNGLGQVGMNTAASD--NFOLSGGQGFAP------IG 219
Db 191 QTDASINHGSGGALVNSLSELMGINTLSPDKNDGHTPEGIGFAPFQLATKIMDKLIR 250
QY 220 QAMAIAGIIRSGGSPVHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLIGISTGCVI 279
Db 251 DGRVIRYIGIGGR-----EIAPLHAQGGI--DQLQGIIVNEVSPDGPANAGIQVNDLI 304
QY 280 TAVDGPAINSATAMADALNGHHPGDVISTWQKSGGTRTGNVTLARGP 328
Db 305 ISVNDKPAISALETMQVAIRPGSVIPV-VNRDDKQLTLQVTIQBY 352

RESULT 4
HTRA_LACLA
ID HTRA_LACLA STANDARD; PRT; 408 AA.
AC Q9L0A6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-) (Htra11).
GN HTRA OR L2136.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN
SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=IL1403;
RX MEDLINE=20177820; PubMed=10712686;
RA Pognet I., Saint V., Seznec E., Simoes N., Bolotin A., Gruss A.;
RT "Htra is the unique surface housekeeping protease in Lactococcus
RT lactis and is required for natural protein processing."
RL Mol. Microbiol. 35:1042-1051(2000).
RN
SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Mauger S., Jaillon O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC
CC -!- FUNCTION: Degrades abnormal exported proteins. Needed for the pro-
CC PEPTIDE PROCESSING OF A NATURAL PRO-PROTEIN AND FOR MATURATION OF
CC A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED
CC PROTEINS.
CC
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AF155705; AAF61294.1; -
CC EMBL; AE006442; AAK06234.1; -
CC PIR; H86891; H86891.
DR MEROPS; S01.273; -
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ-
DR InterPro; IPR001254; Peptidase_S1.

DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydroxylase; Serine protease; Transmembrane; Complete proteome.
FT TRANSMEM 6 26
FT DOMAIN 88 284
FT ACT_SITE 127 127
FT ACT_SITE 157 157
FT ACT_SITE 239 239
SQ SEQUENCE 408 AA; 41648 MW; 581B90B55A7DF851 CRC64;
Query Match 18.6%; Score 315; DB 1; Length 408;
Best Local Similarity 31.6%; Pred. No. 3.6e-11;
Matches 94; Conservative 51; Mismatches 118; Indels 34; Gaps 11;
QY 54 GTGIVIDPNG---VVLNNHVIAGATDINAFSGGQTYGVVGVYDRTQDVAVLQIRG- 109
Db 108 GSGVYKKSGGDYVVTNYHYIAGNSSLDVLLSG-GQVKASVVGYDEYTDLAVLKISSE 166
QY 110 -AGGLPSAAIGGVAVGPPVAVMGNSGGQGTTPRAVGRVVALGQTVQASDSLTAABETL 168
Db 167 HVKDVATFADSSKLTIGEPALVAGSPGLSQPANTATGILSATSRQVLTQGE-NGQTNI 225
QY 169 NGLIQFDDAAIQDGGSGVGVNGLGQVGMNTA---ASDNFOLSGGQGFAPFQIQAAMAI 225
Db 226 NA-IQTDAAINPGNSGGALINIEGVIQTOSKITTEDGTSVEGLGFAIPSDVNNII 284
QY 226 QGIRSGGSPVHIGPTAFGLGVVD-----NNGN-----GARQVRVGSAPAA 269
Db 285 NKLEADG-----KISRPA-LGIRVVDLSQLSTNDSSQKLPSSVTGGVVVSVSGSLPAA 338
QY 270 SLGISTGDVITAVDGPAINSATAMADALNGHHPGDVISTWQKSGGTRTGNVTLAE 326
Db 339 SAGLKAGDVITKVGDTAVTSSTDLSALYSHNINDTVKVTY-RDGKSNATADVLSK 394
RESULT 5
DEGP BRUME
ID DEGP BRUME STANDARD; PRT; 513 AA.
AC Q8VG32;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR BME11330.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN
SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Hasekorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL: AE009571; AAL52511.1; --
PIR: AD3418; AD3418.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_SIC.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SMC0228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR Hydrolase; Serine protease; Periplasmic; Repeat; signal;
KW Complete proteome.
DR SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 513 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 125 299 CATALYTIC.
FT DOMAIN 300 391 PDZ 1.
FT DOMAIN 414 500 PDZ 2.
FT ACT_SITE 152 182 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 182 152 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 513 AA; 53514 MW; C472FEF99DFC6268 CRC64;

Query Match 18.2%; Score 308.5; DB 1; Length 513;
Best Local Similarity 34.3%; Pred. No. le-10;
Matches 104; Conservative 36; Mismatches 118; Indels 45; Gaps 13;

QY 46 GYNNAVGAGTGVIDPNGVLTNNHVTAGATDINAFSV--GSGQTVGVYDVVGYDRTQDVA 103
DB 128 GHERPVAQSGSFVISEDGYVNNHVV--SDGDAYTVLDDGTDLDAKLIGADPRDILA 184
QY 104 VLQIRGAGGLPS-----AATG--GGVAVGPVVMANGSGGGTTPRAVPRVVALGQTQV 156
DB 185 VLKINA----PKRFVYVAFGDDNKVRGDMVAVGNPFGLGST--VTSGLVSARGRDIG 238
QY 157 ASDSLTGAETLNGLIQFDAAIQPDGSGVPVNLGLGVGMNTAASDNFQLSQGGQ--F 214
DB 239 AG-----PYDDFIQIDIAVKNKSGGPAFDLSGEVIGINTAI---FSPSGSGVGIAP 287
QY 215 AIPIGQAMAIAGQIRSGG---GSPTVHIGP-----TAFGLGLGVVDNNGGARVQRVVG 265
DB 288 AIPSTAKQVVDQLIKKGSVERGWIGVQIQVTKDIAASLGLA---EEKGALVASPQDD 343
QY 266 APAASLGISGTVITAYDGAIPINSATAMADALNGHHFPGDVISVTWQKSGGTRTGNVTLA 325
DB 344 GPAAKAGIKAGDVITAVNGETVQDPRDLARKVANIAPGEKAALVWRKNKAEI-NVTIA 402
QY 326 EGP 328
DB 403 AMP 405

RESULT 6
ID DEGP BRUSU STANDARD; PRT; 513 AA.
AC Q44597;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR BRO611.
OS Brucella suis, and
OC Brucella abortus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461, 235;
RN [1]
RC SPECIES=B.abortus; STRAIN=2308;
R SEQUENCE FROM N.A.


```

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
CC SPECIFICITY WITH HHOR/DSGQ.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54548; CAA38420.1; -.
CC EMBL; AEC08704; AAL19173.1; -.
CC FIR; S15337; S15337.
CC MEROPS; S01.273; -.
CC StyGene; SG10173; degP.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 2.
CC Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
KW Complete proteome.
KW SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 475 PROTEASE DO.
FT DOMAIN 281 372 PDZ 1.
FT DOMAIN 378 467 PDZ 2.
FT ACT_SITE 132 132 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 162 162 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 237 237 CHARGE RELAY SYSTEM (POTENTIAL).
FT DISULFID 84 96 BY SIMILARITY.
SQ SEQUENCE 475 AA; 49315 MW; 86E685BF3C1A289F CRC64;

Query Match 17.4%; Score 296; DB 1; Length 475;
Best Local Similarity 32.2%; Pred.No. 4.9e-10;
Matches 96; Conservative 46; Mismatches 102; Indels 54; Gaps 12;

Qy 54 GTGIVID-PNGVVLNNHVIAGATDINAFVSGSQTYGVDVYGDRTQDVAVLQLRGAGG 112
Db 115 GSGVIIDAAKGYYVNNHVDNASVIKV-QLSDGRKFDKAVGKDPKSDIALIQNPKN 173
Qy 113 LPSAIOGGVA--VGEPPVAMNGSGGGGTTPRAVPGRVVALGQTVQASDSLTGAE-ETLN 169
Db 174 LTAIKLADSDALRYGDYTVAGNPFPGIGET--VTSGIVSALGR-----SGLNVNVE 223
Qy 170 GLTIQDAAIQDGGGPPVWNLGQWGMNTA--ASDNFQLSQGGQGAIP-----ICQA 221
Db 224 NFIQTDAIRNGSGGALVNLNGELIGINTAILAPDGGNI---GIGFAIPSNVYKNLTQS 280
Qy 222 MAIAGQIRSGGSPETHIGTAFILGLGVDDNNGH-----GARVQRVVGGA 268
Db 281 MVVEYGVKRG-----ELGIMGTELNSELAKAMKVDQAQGFVSQVMPNSSA 326
Qy 269 ASLGISGTGDVITAVDGAIPINSATAMADALNGHHHPGDVTSVTWQTKSGGTRGNVTLAE 326
Db 327 AKAGIKAGDVITSLNGKRFPSFAALRAQVGTMFVGSKISL-GLIREGKAITVNLLELQ 393

RESULT 9

```

RESULT 9

FT CONFLICT 464 504 KTVTSPTTSSPART (IN REF. 1).
 FT BEST LOCAL SIMILARITY 33.7%; Score 296; DB 1; Length 504;
 SQ SEQUENCE 504 AA; 53035 MW; D7E82B9981EA23C CRC64;
 Matches 98; Conservative 39; Mismatches 116; Indels 38; Gaps 12;
 Query Match
 54 GTGVIDNGVLLNNHVIAGATDINAFSV--GGQTYGVVDVYDRTQDAVLQL--RG 109
 DB 124 GSGFFITDGLVNTNNHV---SGSAFTVIMNDGTDLAKLVGKDSRTDLAVLKVDK 180
 QY 110 AGGLPSAIGGVAVGPEVVMGNSGGGTGPRVPRVVALGQTVQASLSLTCAEETLN 169
 DB 181 KFTVVSFADDEKVRVGVVAVGPNFGLGGVTVA--GLISARGRDIGSG-----PYD 230
 QY 170 GLIQDAAIQDGGPVGNGVGWGMNTAASDNFQSQGGG--FAIFIGQAMAIAGQ 227
 DB 231 DYLQVDAVNRGNSGGPTFNLGSGWGINTA---FPFSGGVGIAFAIPASVAKDVWDS 287
 QY 228 IRSGG---GSPVTHIGFTA-----FLGLGVDMNGNGARVQVRVGSAPASLIGISTGV 278
 DB 288 LIKDGTVSRGMLGVQIQFVTKDIAESLGL-----SEANGALVWFPQAGSGEKAGIKNGDV 343
 QY 279 ITAVDGAIPINSATAMADALNHHPGDDVISVT--WOTKSGGTRTGNVTLAEGP 328
 DB 344 VTALNGEPVDPRLARRVAALRGSTAEVTLW---RSKSETVNLGIGTLP 392

RESULT 10
 HTRA_BACSU STANDARD; PRT; 449 AA.
 AC O34358;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 GN HTRA OR BSUI2900.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Devine K.M.;
 RL "Sequence of the Bacillus subtilis genome between xlyA and ykoR.";
 RN Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bextero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brullet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrar J., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizi A., Galleron N.,
 RA Gim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Juris B., Karamata D., Kasahara Y., Klaere-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ozawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Pressac E., Pujić A., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 RN [3]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=20158875; PubMed=10692364;
 RA Noone D., Howell A., Devine K.M.;
 RT "Expression of ykdA, encoding a Bacillus subtilis homologue of HtrA,
 RT is heat shock inducible and negatively autoregulated.";
 RL J. Bacteriol. 182:1592-1599(2000).
 RN [4]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=20576168; PubMed=11133960;
 RA Noone D., Howell A., Collier R., Devine K.M.;
 RT "ykdA and yvta, HtrA-like serine proteases in Bacillus subtilis,
 RT engage in negative autoregulation and reciprocal cross-regulation of
 RT ykdA and yvta gene expression.";
 RL J. Bacteriol. 183:654-663(2001).
 RN [5]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=21439741; PubMed=1155295;
 RA Hytynlaenen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
 RA Vitikainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M.,
 RA Kontinen V.P.;
 RT "A novel two-component regulatory system in Bacillus subtilis for the
 RT survival of severe secretion stress.";
 RL Mol. Microbiol. 41:1159-1172(2001).
 CC -!- FUNCTION: May be involved in processing, maturation, or secretion
 CC of extracellular enzymes.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -!- INDUCTION: Transcription is cssS dependent. Induced by heat shock
 CC during exponential growth and by heterologous amyases at the
 CC transition phase of the growth cycle. Negatively regulates its own
 CC expression during exponential growth and during heat shock.
 CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
 CC of yvta, especially during stress conditions.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ002571; CAA05570.1; -;
 CC EMBL; Z91110; CAB13147.1; -;
 CC PIR; A69643; A69643.
 CC MEROPS; S01.273; -;
 CC Subtilist; B012608; htra.
 CC InterPro; IPR005003; Cys_Ser_trypsin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001940; Peptidase_S1C.
 CC Pfam; PF00595; PDZ; 1.
 CC Pfam; PF00089; trypsin; 1.
 CC PRINTS; PR00834; PROTEASES2C.
 CC SMART; SM00228; PDZ; 1.
 CC PROSITE; PS0106; PDZ; 1.
 CC Hydrolyase; Protease; Serine protease; Heat shock; Transmembrane;
 KW Complete proteome.

FT	DOMAIN	1	44	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	45	67	POTENTIAL.
FT	DOMAIN	68	439	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	103	108	POLY-SER.
FT	DOMAIN	146	152	POLY-SER.
FT	DOMAIN	348	437	PDZ.
FT	ACT SITE	179	179	CHARGE RELAY SYSTEM (POTENTIAL).
FT	ACT SITE	209	209	CHARGE RELAY SYSTEM (POTENTIAL).
FT	ACT SITE	290	290	CHARGE RELAY SYSTEM (POTENTIAL).
SQ	SEQUENCE	449 AA;	47713 MW;	E12B07A9018EE414 CRC64;
Query Match				
Best Local Similarity 29.3%; Score 294.5; DB 1; Length 449;				
Matches 93; Conservative 52; Mismatches 127; Indels 45; Gaps 11;				
QY	28	SAMVQGVQVNVNINIKLGYNNA-----VGAGTGIVL-DPNG--VVLNN	69	
DB	119	SDMVEDLSAIVGIVITMLQXSSSLFGSSSSDSTESGSGVIFKKGKAVIINN	178	
QY	70	HVIAGATDINAFSVSGQTYGVVDVGYDRTQDAVLQLRGAGLPSAIG--GGVAVGEP	127	
DB	179	HVVEGASSLKV-SLYDGTETVAKLVGSDSLTDLAVLIQISDDHVTKVANFGDSSDLRTGET	237	
QY	128	VVAMGNSGGGTPRAVGRVVALGQVQASDLSLTGAETLNGLIQFDAAIOFGDSGGPV	187	
DB	238	VIAIGDPLGKDLISRTVTQGVISGVDTVSMST--TSAGETSNVITQDAAINPGSGGGL	294	
QY	188	VNLGLGWGNT--AASDNFQLSQGGQGAIPFGQAMAIAGQIRSGGSGFTVHIGTAPL	245	
DB	295	LNTDGKIVGINSKISDEDDVE----GIGFAIPNDVKPIABELLSKGQIERVIG-VSM	349	
QY	246	GLGVVDNN-----NGARVQRVVGSAAPASIGSTGVITAVDAPINSATA	292	
DB	350	DLEQVPQNVQEGTGLFGSLNKGVYIREVAGSPAERKAGLAEIIIGLKGEIDTSGE	409	
QY	293	MADAL-NGHHPGDVISV	308	
DB	410	LRNLYKDAKIGDTVEV	426	
RESULT 11				
ID	DEGP_ECOLI	STANDARD;	PRT;	474 AA.
AC	P09376; E15724;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DE	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Protease do precursor (EC 3.4.21.-)			
GN	DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.			
OS	Escherichia coli, and			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562, 83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RC	MEDLINE=90507448; PubMed=3057437;			
RA	Lipinska B., Sharma S., Georgopoulos C.;			
RA	"Sequence analysis and regulation of the htrA gene of Escherichia			
RT	coli: a sigma 32-independent mechanism of heat-inducible			
RT	transcription."			
RL	Nucleic Acids Res. 16:10053-10067(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / W3110;			
RC	MEDLINE=94261430; PubMed=8202364;			
RA	Fujita N., Mori H., Yura T., Ishihama A.;			
RA	"Systematic sequencing of the Escherichia coli genome: analysis of			
RT	the 2.4-4.1 min (110,917-193,643 bp) region.;"			
RL	Nucleic Acids Res. 22:1637-1639(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			

RC	STRAIN=K12 / MG1655;			
RC	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Ferna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.;"			
RL	Science 277:1453-1474(1997).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,			
RA	Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,			
RA	Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,			
RA	Davis R.W.;			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RC	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA	Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grobbeck B.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,			
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Weich R.A., Blattner F.R.;			
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"			
RL	Nature 409:529-533(2001).			
RN	[6]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / RIMD 0509952;			
RC	MEDLINE=21156231; PubMed=11258796;			
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA	Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,			
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli			
RT	O157:H7 and genomic comparison with a laboratory strain K-12.;"			
RL	DNA Res. 8:11-22(2001).			
RN	[7]			
RP	SEQUENCE OF 1-50 FROM N.A.			
RC	STRAIN=K12;			
RC	MEDLINE=90323597; PubMed=2165018;			
RA	Quirk S., Bhatnagar S.K., Bessman M.J.;			
RT	"Primary structure of the deoxyguanosine triphosphate			
RT	triphosphohydrolase-encoding gene (dgt) of Escherichia coli.;"			
RL	Gene 89:13-18(1990).			
RN	[8]			
RP	SEQUENCE OF 1-16 FROM N.A.			
RC	MEDLINE=90207273; PubMed=2157212;			
RA	Wurgler S.M., Richardson C.C.;			
RT	"Structure and regulation of the gene for dGTP triphosphohydrolase			
RT	from Escherichia coli.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).			
RN	[9]			
RP	SEQUENCE OF 27-39, AND CHARACTERIZATION.			
RC	MEDLINE=90202693; PubMed=2180903;			
RA	Lipinska B., Zylcz M., Georgopoulos C.;			
RT	"The HtrA (DegP) protein, essential for Escherichia coli survival at			
RT	high temperatures, is an endopeptidase.;"			
RL	J. Bacteriol. 172:1791-1797(1990).			
RN	[10]			
RP	SEQUENCE OF 27-30.			
RC	STRAIN=K12 / W3110;			
RC	MEDLINE=98263247; PubMed=9600841;			
RA	Wilkins M.R., Gasteiger E., Tonelli L., Ou K., Tyler M.,			
RA	Sanchez J.-C., Gooley A.A., Walsh B.J., Bairoch A., Appel R.D.,			
RA	Williams K.L., Hochstrasser D.F.;			
RT	"Protein identification with N and C-terminal sequence tags in			
RT	proteome projects.;"			
RL	J. Mol. Biol. 278:599-608(1998).			
RN	[11]			
RP	IDENTITY OF HTRA AND PROTEASE DO.			
RP	MEDLINE=9122240; PubMed=2025286;			

Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
 Tanaka K., Ichihara A., Ha D.B., Chung C.H.,
 "Protease do is essential for survival of *Escherichia coli* at high
 temperatures: its identity with the *htrA* gene product.",
 RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
 RN (12)
 RP DISULFIDE BOND.
 RX MEDLINE=22760311; PubMed=12878036;
 RA Skorko-Glonek J., Zurawa D., Tanfani F., Seire A., Mawrynow A.,
 RA Nakiewicz J., Bertoli E., Lipinska B.,
 RT "The N-terminal region of *HtrA* heat shock protease from *Escherichia*
 RT coli is essential for stabilization of *HtrA* primary structure and
 RT maintaining of its oligomeric structure.",
 RL Biochim. Biophys. Acta 1649:171-182(2003).
 CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE
 CC ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGO.
 CC -!- SUBUNIT: Multimeric.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- INDUCTION: By heat shock.
 CC -!- MISCELLANEOUS: *HtrA* IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
 CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M36536; AAA23994.1; -;
 DR EMBL; X12457; CAA30997.1; -;
 DR EMBL; D26562; BAB98738.1; -;
 DR EMBL; AE000125; AAC73272.1; -;
 DR EMBL; U70214; AAB08591.1; -;
 DR EMBL; AE005192; AAG54465.1; -;
 DR EMBL; AP002550; BAB33588.1; -;
 DR EMBL; M29955; AAA23717.1; -;
 DR EMBL; M31772; AAA23680.1; -;
 DR PIR; B85500; E85500.
 DR PIR; E90849; E90649.
 DR PIR; S45229; S45229.
 DR PDB; 1KY9; 03-APR-02.
 DR MEROPS; S01.273; -;
 DR SWISS-2DPAGE; P09376; COLI.
 DR EcoGene; EG10463; degP.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; tryptsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 KW Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; signal;
 KW Complete proteome; 3D-structure.
 FT SIGNAL 1 26
 FT CHAIN 27 474
 FT DOMAIN 280 371
 FT PDZ 1.
 FT PDZ 2.
 FT ACT_SITE 131
 FT ACT_SITE 161
 FT ACT_SITE 236
 FT ACT_SITE 236
 FT DISULFID 83 95
 FT CONFLICT 10 10
 FT CONFLICT 46 46
 FT CONFLICT 192 192
 FT CONFLICT 467 474
 FT CONFLICT 467 474
 FT REF. 1).
 FT STYLDQM -> RHLFVNAVSLNPLKTRGSPNVL (IN
 FT REF. 1).
 FT

SQ SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;
 Query Match 17.3%; Score 294.5; DB 1; Length 474;
 Best Local Similarity 33.6%; Pred. No. 5.9e-10;
 Matches 89; Conservative 42; Mismatches 81; Indels 53; Gaps 11;
 QY 54 CTGIVIDPN-GVLTNNHVIAGATINAFSVSGGTGYGVVGVYDRTQDVAVLQRGAGG 112
 DB 114 GSGVIADKGVVYVNNHVDNATVIKV-QLSDGRKFDKMKVGDPRSDIALIQINPKN 172
 QY 113 LPS--AAIGGGVAVGEPVVMGNSGGQGTTPRAVPRVVALGQTVQASDSLTGAB-ETLN 169
 DB 173 LTAIKWADSDALRVGDYTVAGNPFGLGET--VTSGIVSALGR-----SGLNAENYE 222
 QY 170 GLIQDAAIQPDGSGPVVNGLGQVVGNTA--ASDNFQLSQGGGQPAIP-----IQQA 221
 DB 223 NFQIQTDAAINRNGSGGALVNLNGELIGINTAILAPDGGNI---GIGFAIPSNMVKNLTSQ 279
 QY 222 MAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGN-----GARYQRVVGSAPA 268
 DB 280 MVEYGVQVREG-----ELGIMGTLSLAKAMKVDAAQAGFVSQVLPNSSA 325
 QY 269 ASLGISTGVTAVDGPINSATAM 293
 DB 326 AKAGIKAGDVITSLNGKPISSPAAL 350
 RESULT 12
 HTQA_HAEIN
 ID HTQA_HAEIN STANDARD; PRT; 466 AA.
 AC P45129;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable periplasmic serine protease do/hhoA-like precursor
 DE (EC 3.4.21.-).
 GN H11259.
 OS *Haemophilus influenzae*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; *Haemophilus*.
 OX NCBI_TaxID=727;
 RN [1]_TaxID=727;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
 RA Fire L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of *Haemophilus influenzae*
 RT Rd.",
 RL Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (potential).
 CC -!- SIMILARITY: SEEMS TO BE AN INTERMEDIATE FORMS BETWEEN E.COLI HTRA
 CC (PROTEASE DO) AND HHOA.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U32805; AAC22906.1; -;
 DR PIR; A64113; A64113.

Query Match	16.9%;	Score 287;	DB 1;	Length 478;
Best Local Similarity	30.9%;	Pred. No. 1.6e-09;		
Matches	89;	Conservative 50;	Mismatches 109;	Indels 40; Gaps 12;

QY	54	GTGIVIDPN-GVVLNNHVIAGATDINAFSGSGQTYGVVDYDRTQDVAVLQLRGAGG	112
Db	116	GGVLIINADKGYAVTNNHVENANKIQV-QLSDGRYEAREVIGKDSRSDIALIQLKNANN	174
QY	113	LPSAAI--GGGVAVGEPVVAMNSGGQGTPEAVFGRVVVALGQTVQASDSLTGAE-ETLN	169
Db	175	LSEIKIADSDNLRVGDYTVAINPYGLGET--VTSGIISALGR-----SGLNIEHYE	224
QY	170	GLIOPDAIQCDSGSPVNG-GQVVGMNTA--ASDNFOLSGOGGPAIPICQAMAIAGQ	227
Db	225	NFIQTDAALNRGNSGALVNLKGLIGINTAILAPDGGNI---GIGFALPCNMVKNLTAQ	281
QY	228	IRSGGGSPTVHIGPTAFLGLGVVD-----NNGNGARVQRVWGSAPAAISLGISTGDVI	279
Db	282	MTQFG--QVRRGELGIMGMELNSDLAQIMKINSQKGFVSRVLPNSSAFEAGIKAGDII	338
QY	280	TAVDGAPINSATAMADALNGHHFGDVISVTWTKSGGTGTG---NVTL	324
Db	339	ISLNRKPISSFSLSRAEIGS-----LPVATKMWELGVFREGRIKNITV	380

Search completed: June 22, 2004, 17:16:47
Job time : 6.58295 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 26.9826 Seconds
(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSQDRADF.....QTKSGGTRGNTVLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315318202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25:*
- 1: sp_archea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_plant:*
 - 10: sp_rodent:*
 - 11: sp_virus:*
 - 12: sp_vertebrate:*
 - 13: sp_unclassified:*
 - 14: sp_virus:*
 - 15: sp_bacteriaph:*
 - 16: sp_bacteriaph:*
 - 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	1645	96.9	355	16	007175	007175 mycobacteri
2	1645	96.9	355	16	Q7U289	Q7U289 mycobacteri
3	1231.5	72.5	361	2	Q50320	Q50320 mycobacteri
4	1198	70.6	354	16	Q9CCY9	Q9CCY9 mycobacteri
5	460.5	27.1	446	16	Q8VKA4	Q8VKA4 mycobacteri
6	460.5	27.1	464	16	Q7U0X2	Q7U0X2 mycobacteri
7	459.5	27.1	464	16	O53896	O53896 mycobacteri
8	450	26.5	382	16	Q9CD67	Q9CD67 mycobacteri
9	450	26.5	452	2	Q2Z566	Q2Z566 mycobacteri
10	397.5	23.4	519	16	Q93J30	Q93J30 streptomyc
11	396	23.3	375	16	Q8DG87	Q8DG87 synchococc
12	393.5	23.2	542	16	Q9FBK9	Q9FBK9 streptomyc
13	384	22.6	472	16	Q82FM9	Q82FM9 streptomyc
14	384	22.6	619	16	Q82I18	Q82I18 streptomyc
15	379	22.3	473	16	Q8F817	Q8F817 corynebact
16	374	22.0	394	16	P72780	P72780 synchocyst

17	368	21.7	407	16	Q8YTF9	Q8YTF9 anabaena sp
18	362	21.3	326	16	Q8DMV9	Q8DMV9 synchococc
19	357.5	21.1	675	16	Q8G6T3	Q8G6T3 bifidobacte
20	357	21.0	413	16	O8NS10	O8NS10 corynebacte
21	354.5	20.9	395	16	Q7V5C8	Q7V5C8 prechloroco
22	353.5	20.8	391	16	Q899T5	Q899T5 clostridium
23	353	20.8	433	16	Q37GB5	Q37GB5 clostridium
24	353	20.8	447	16	Q8R756	Q8R756 thermococci
25	352	20.7	452	16	P73354	P73354 synchocyst
26	351	20.7	339	16	Q89RP2	Q89RP2 bradyrhizob
27	350	20.6	525	16	Q8PMV4	Q8PMV4 xanthomonas
28	349	20.6	500	2	Q9KJN6	Q9KJN6 myxococcus
29	348	20.5	374	16	Q7U435	Q7U435 synchococc
30	348	20.5	490	16	Q8XP75	Q8XP75 ralsstonia s
31	346.5	20.4	389	16	Q8DL28	Q8DL28 synchococc
32	346	20.4	371	16	O31388	O31388 bradyrhizob
33	345.5	20.3	525	16	Q89QJ8	Q89QJ8 bradyrhizob
34	343.5	20.2	432	16	Q7U8K9	Q7U8K9 synchococc
35	339	20.0	514	16	Q87C10	Q87C10 xylella fas
36	338	19.9	514	16	Q9PBA3	Q9PBA3 xylella fas
37	336.5	19.8	377	16	Q7U604	Q7U604 synchococc
38	335.5	19.8	362	16	Q9S2K5	Q9S2K5 streptomyc
39	335.5	19.8	465	16	Q92QB6	Q92QB6 rhizobium m
40	335	19.7	511	2	Q8RTK2	Q8RTK2 xanthomonas
41	335	19.7	525	16	Q8PB56	Q8PB56 xanthomonas
42	333	19.6	416	16	Q8YVH0	Q8YVH0 anabaena sp
43	332	19.6	398	16	O83752	O83752 treponema p
44	332	19.6	441	16	Q9RTK4	Q9RTK4 deinococcus
45	331	19.5	371	16	Q7VA24	Q7VA24 prochlorococ

ALIGNMENTS

RESULT 1

007175 ID 007175 PRELIMINARY; PRT; 355 AA.

AC 007175; (TREMELrel. 04, Created)

DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Hypothetical protein (Serine protease, putative).

GN PEPA OR RV0125 OR MTC1418B.07 OR MT0133.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence."

RN [2]

RX Nature 393:537-544 (1998).

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J.F., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Ormrod L.A., Remington K., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains."

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

```

CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; Z96071; CAB09453.1; -.
DR EMBL; AE0006925; AAK44357.1; -.
DR PIR; F70983; F70983.
DR TIGR; MT0133; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR008256; Peptidase_S1b_v8.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Hypothetical protein; Serine protease; Protease;
KW Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 96.9%; Score 1645; DB 16; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.4e-81;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNITKLGYNNAVAGTGVIDPNGVVL 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNITKLGYNNAVAGTGVIDPNGVVL 92
QY 68 NNHVIAGATDINAFSVSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVMGNSGGQGTTPRAVGRVVALQGTVOASDSLGTGAETLNGLIQFDAAIQPDSDSGPV 187
DB 153 VVMGNSGGQGTTPRAVGRVVALQGTVOASDSLGTGAETLNGLIQFDAAIQPDSDSGPV 212
QY 188 VVGLGVVGMNTAASDNFQSGGQGFALPIQOAVAIAGQIRSGGSGPTVHIGTAFGL 247
DB 213 VVGLGVVGMNTAASDNFQSGGQGFALPIQOAVAIAGQIRSGGSGPTVHIGTAFGL 272
QY 248 GVVDNNGGARVQVVGSAAPASLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGGARVQVVGSAAPASLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQKSGGTRTGNVTLAEGPPA 355

RESULT 3
Q50320 PRELIMINARY; PRT; 361 AA.
ID Q50320 PRELIMINARY; PRT; 361 AA.
AC Q50320; 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE 34kDa protein precursor.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JD88/107;
RX MEDLINE=95005449; PubMed=7921248;
RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterisation of a putative serine protease
expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
RL Microbiology 140:1977-1982(1994).
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; Z23092; CAA80638.1; -.
DR PIR; S47170; S47170.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1c.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.

```


KW Hydrolase; Protease; Serine protease; Signal.
 FT SIGNAL 1 38 POTENTIAL.
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 72.5%; Score 1231.5; DB 2; Length 361;
 Best Local Similarity 73.1%; Pred. No. 4.9e-59;
 Matches 236; Conservative 38; Mismatches 48; Indels 1; Gaps 1;

QY 8 APPALSQDRFAFFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGVIDPNGVLT 67
 DB 40 ARSGLALDFADRLAPIDPSAMVQGVGVVNTKFGYNNAVAGAGTGVIDPNGVLT 99
 QY 68 NNHVIAGATDINAFSVGSQTVGVVGVVNTKLGYNNAVAGAGTGVIDPNGVLT 127
 DB 100 NNEVISGATEISAFDVGNGQTVADVGVVNTKLGYNNAVAGAGTGVIDPNGVLT 159
 QY 128 VVAMGNSGGGTPPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFOAAIOPGSGGVP 187
 DB 160 IVALNVGGGGTPPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFOAAIOPGSGGVP 219
 QY 188 VNLGQVGVGNTAASDNFOLSGGGQFAIPGOAMAIAGQIRSGGSPVTHIGPTAFGL 247
 DB 220 VNSAGQVIGVDTAATDSYKMS-GGGQFAIPIGRAMAVANQIRSGAGSNVTHIGPTAFGL 278
 QY 248 GVVDNNGGARVORVVGSAASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVTS 307
 DB 279 GVVDNNGGARVORVVGSAASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVTS 338
 QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
 DB 339 VHFPSVDGGERTANITLAEAGPPA 361

RESULT 4
 Q9CCY9 PRELIMINARY; PRT; 354 AA.
 ID AC Q9CCY9
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Probable secreted serine protease.
 GN ML2659.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 CC -/- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL583926; CAC32191.1; -;
 DR PIR; A87242; A87242.
 DR Leproma; ML2659; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR003256; Peptidase_S1B.
 DR InterPro; IPR001940; Peptidase_S1C.

QY 9 PPALSQDRFAFFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGVIDPNGVLT 68
 DB 37 PSTLALDRFSRRPLPLNPAWVA---PQVNVSTRLGYNNAVAGAGTGVIDSSGVLT 93
 QY 69 NNHVIAGATDINAFSVGSQTVGVVGVVNTKLGYNNAVAGAGTGVIDPNGVLT 128
 DB 94 NNHVIAGATDINAFSVGSQTVGVVGVVNTKLGYNNAVAGAGTGVIDPNGVLT 153
 QY 129 VVAMGNSGGGTPPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFOAAIOPGSGGVP 188
 DB 154 VALNGTGGGGTPPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFOAAIOPGSGGVP 213
 QY 189 NGLGQVGVGNTAASDNFOLSGGGQFAIPGOAMAIAGQIRSGGSPVTHIGPTAFGL 248
 DB 214 NSRGQVGVGNTAASDNFOLSGGGQFAIPGOAMAIAGQIRSGGSPVTHIGPTAFGL 272
 QY 249 VVDNNGGARVORVVGSAASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVTS 308
 DB 273 VLDNNGGARVORVVGSAASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVTS 332
 QY 309 TWQKSGGTRTGNVTLAEGPPA 330
 DB 333 NYRSAGGDLTANVTLAEGPPA 354

RESULT 5
 Q8VK44 PRELIMINARY; PRT; 446 AA.
 ID AC Q8VK44
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Heat shock protein Htra, putative.
 GN MT1011.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeSoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE006985; AAK45259.1; -;
 DR TIGR; MT1011; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR SEQUENCE 446 AA; 44484 MW; 54170CBEA8F872B CRC64;

Query Match 27.1%; Score 460.5; DB 16; Length 446;
 Best Local Similarity 37.4%; Pred. No. 2.5e-17;
 Matches 122; Conservative 55; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGPQVNNIKLGYNNVAGAGTGVIDPNGVLTNNHVIAGAT- 76
 DB 129 PAANMPGSEVQVAAKVPVSVVMTDLGRQSE--EGSGILSAEGLITNNHVIATAAK 186

QY 77 -----DINAFVSGSQTVGVYDRTQDVAVLQRLGAGGLPSAAIGGG--VAVGEP 127
 DB 187 PPLGSPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVQVSGSLTPISLGSSDDLVRVGP 246

QY 128 VVAMGNSGGQGTTPRAVGRVVALGQTVQASDSLGTGAETLNGLIQDAAIQPDGSGGPV 187
 DB 247 VLAIGSPGLEGT--VTTGIVSALNRPVSTTGE--AGNQNTVLDATQDAAINPNSGGAL 303

QY 188 VNLGLQVVGMMNTA---ASDNFQLSQG--GQGFAPICQAMAIAGQIRSGGSGPTVHIGP 241
 DB 304 VNNNAQLGVNSAIATLGAADSADAQSGSIGLGFAPVQAKRIADELISTGKA-----S 357

QY 242 TAFGLGLGVV-DNNGNGARVORVWGSAPASLSIGTSDVITAVDGPAPINSATMADALNGH 300
 DB 358 HASLGVQVNTDKDTPGAKIVEVAGGAANAAGVPGVVTVKVDRPINSADALVAARSK 417

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAE 326
 DB 418 AFGATVALTQDPGSGSRTVQVTLCK 443

RESULT 6
 QY002 ID Q7U0X2 PRELIMINARY; PRT; 464 AA.
 AC Q7U0X2;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Probable serine protease (Serine proteinase) (EC 3.4.21.-).
 GN MB1009.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248337; CAD93870.1; -.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 464 AA; 46436 MW; AE93A4BB3FFA9BE3 CRC64;

Query Match 27.4%; Score 460.5; DB 16; Length 464;
 Best Local Similarity 37.4%; Pred. No. 2.6e-17;
 Matches 122; Conservative 55; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGPQVNNIKLGYNNVAGAGTGVIDPNGVLTNNHVIAGAT- 76
 DB 147 PAANMPGSEVQVAAKVPVSVVMTDLGRQSE--EGSGILSAEGLITNNHVIATAAK 204

QY 77 -----DINAFVSGSQTVGVYDRTQDVAVLQRLGAGGLPSAAIGGG--VAVGEP 127
 DB 205 PPLGSPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVQVSGSLTPISLGSSDDLVRVGP 264

QY 128 VVAMGNSGGQGTTPRAVGRVVALGQTVQASDSLGTGAETLNGLIQDAAIQPDGSGGPV 187
 DB 265 VLAIGSPGLEGT--VTTGIVSALNRPVSTTGE--AGNQNTVLDATQDAAINPNSGGAL 321

QY 188 VNLGLQVVGMMNTA---ASDNFQLSQG--GQGFAPICQAMAIAGQIRSGGSGPTVHIGP 241
 DB 322 VNNNAQLGVNSAIATLGAADSADAQSGSIGLGFAPVQAKRIADELISTGKA-----S 375

QY 242 TAFGLGLGVV-DNNGNGARVORVWGSAPASLSIGTSDVITAVDGPAPINSATMADALNGH 300
 DB 376 HASLGVQVNTDKDTPGAKIVEVAGGAANAAGVPGVVTVKVDRPINSADALVAARSK 435

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAE 326
 DB 436 AFGATVALTQDPGSGSRTVQVTLCK 461

RESULT 7
 OS3896 ID OS3896 PRELIMINARY; PRT; 464 AA.
 AC OS3896;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Putative serine protease.
 GN RV0983 OR MT0044.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krohn A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544(1998).
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL021999; CAA17582.1; -.
 DR PIR; C70821; C70821.
 DR Tuberculin; RV0983; -.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR004478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 464 AA; 46452 MW; AB93BFCG53E1EC8F CRC64;

Query Match 27.1%; Score 459.5; DB 16; Length 464;
 Best Local Similarity 37.4%; Pred. No. 2.9e-17;
 Matches 122; Conservative 55; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGPQVNNIKLGYNNVAGAGTGVIDPNGVLTNNHVIAGAT- 76
 DB 147 PAANMPGSEVQVAAKVPVSVVMTDLGRQSE--EGSGILSAEGLITNNHVIATAAK 204

QY 77 -----DINARVSGQTYGVVDVYDRTQDVAVLQLEACAGLPSAIGGG--VAVGEP 127
Db 205 PPLGSPPPKTTVTFSQDRTAPTWTGADFTSDIAVVRVQVSGSLTPISLGSSSLRLRVGP 264
QY 128 VVAMNGSGGGGTPRAVPGRVVAGLQTVQASDSLTGAETLNGLIQFOAAIOPGSGGPPV 187
Db 265 VLAIGSPGLEGT--VTTGIVSALNRPVSTTGE--AGNQNTVLDATQDAANPGNSGGAL 321
QY 188 VNLGLQVVGNTA-----ASDNFQLSG--GQGFAPIGQAMAIAGQIRSGGSPVTHIGP 241
Db 322 VNMNAQLVGVNSAIALTGADSDAQAQSSIGLGFAPVDAQRIADELISTGKA-----S 375
QY 242 TAFGLGVV--DNNNGARVQRVVGSAASLGISTGDVITAVDGAIPNSATAMADALNGH 300
Db 376 HASLGVQVNDKDTLGAKEVVEVAGGAANAGVPGKGVVTVKDDRPINSADALVAARVSK 435
QY 301 HFGDVISVTWQTKSGGTRTGNVTLAE 326
Db 436 ACPATVALTQDPGSGGSRVTQVTLGK 461

RESULT 8

Q9CD67 ID Q9CD67 PRELIMINARY; PRT; 382 AA.
AC Q9CD67
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Possible secreted serine protease.
GN ML0176.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mangall K., Braham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Dutcoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Soares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrall B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583917; CAC29684.1; -
DR PIR; H86930; H86930.
DR Leproma; ML0176; -
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 382 AA; 37084 MW; 3DD8DD8AE32A80D CRC64;

Query Match 26.5%; Score 450; DB 16; Length 382;
Best Local Similarity 37.5%; Pred. No. 7.5e-17;
Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;
18 ADPFPALPDSMAVQVQ----PQVNVNTKLGYNNAVGAGTGIVIDPNGWLTNNHVA 73

Db 59 ASVPAANM-FSGSVEQVAVKVVSVWLETDLGRQSE--EGSGVILSADGLILTNHVA 115
QY 74 GA-----TDINAFVSGSQTYGVVDVYDRTQDVAVLQLEACAGLPSAIG 119
Db 116 VAAKPGGGPGGLSPKTTVTFF---DGRASFTVVGADFTSDIAVVRVQVSGSLTPIIMG 172
QY 120 GG--VAVGEPVWAMNGSGGGGTPRAVPGRVVAGLQTVQASDSLTGAETLNGLIQFDAA 177
Db 173 SSADLRVGQPVVAVGSPGLAGT--VTSGIVSALNRPVSTTGE--SGNQNTVLDATQDAA 229
QY 178 IQPGSGGPPVNVGLGVVGMNTAA-----SDNFQLSG--GQGFAPIGQAMAIAGQIRSG 231
Db 230 INPNSGGALVNMGSQLVGVNSAIALTGADSDAQAQSSIGLGFAPVDAQRIADELIST 289
QY 232 CGSTVHIGTAFGLGLGVVDNNGN--GARVQRVVGSAASLGISTGDVITAVDGAIPNSA 290
Db 290 G--KATH-----ASLGVQVATDKGTPGAKVMDVWAGGAANAAPKGVVLTQVDDRLISSA 343
QY 291 TAMADALNGHHPGDVSVVTWQTKSGGTRTGNVTLAE 326
Db 344 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 379

RESULT 9

Q9Z5G6 ID Q9Z5G6 PRELIMINARY; PRT; 452 AA.
AC Q9Z5G6
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Putative serine protease.
GN MLCB373.28.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RL Harris D., Taylor K.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA James K.D., Parkhill J., Barrall B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eigmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae."
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL; AL035500; CAB36690.1; -
DR PIR; T45448; T45448.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 452 AA; 45130 MW; 6CA675EB0911F983 CRC64;

Query Match 26.5%; Score 450; DB 2; Length 452;
Best Local Similarity 37.5%; Pred. No. 9.3e-17;
Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFFPALPLDPSAMVAQVG-----POVNNINTKLYNNVAGAGTGIVDPNGWLTNNHVA 73
 Db 129 ASVPAANM-PSGSVEQAVKVPVSWLTDLGRQSE--EGSGVLSADGLIITNNHVA 185
 QY 74 GA-----TDINAFSGSQGTGVGVVGVVDRDQVAVLQRLGAGGLPSAIG 119
 Db 186 VAAKPGGGGGLSPKTTVTF-----DGTASTFTVVGADPTSDIAVVRVQISGLTITMG 242
 QY 120 GG--VAYGEPVAMGSGGGGTTPRAVPGRVVALGQTVQASDSLGTGAETLNGLIQFDDA 177
 Db 243 SSADLRVGQPVVAVGSPGLAGT--VTSGIVSALNRVPSTTGE-SGNQNTVLDIAITDAA 299
 QY 178 IQPGDSGGPVVGLGQVGVNNTAA-----SDNFQSGGQGAFAIPGQMAIAGIRSG 231
 Db 300 INPNSGGALVNMGGQLVGVNSAIATLGADSGSISGLGFAIPVDQAKRIADELIST 359
 QY 232 GGSPTVHIGPTAFGLGVVNDNGN-GARVQVVGSAAPASLGISTGDTVITAVDGAIPNSA 290
 Db 360 G--KATH-----ASLGQVVAIDKTPGPAKVDVWVAGGAANAAPVKGVLTKVDDRLISSA 413
 QY 291 TAMADALNHHHPGPDVIVSVTWQTKSGGTRTGNVTIAE 326
 Db 414 DALVAARSKAPGDKVSLTYQDQSGSSRTVQVTLGK 449

RESULT 10
 Q93J30 ID Q93J30 PRELIMINARY; PRT; 519 AA.
 AC Q93J30;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC03977 OR SCBAC25E3.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Collins M., Harris D.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RL "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 Mol. Microbiol. 21:77-96(1996).
 [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Horneby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RL "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)";
 Nature 417:141-147(2002).

DR EMBL; AL939118; CAC44701.1; -.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR Hydrolase; Protease; Serine protease; Complete proteome.
 SK SEQUENCE 519 AA; 50327 MW; 5CB8D5F0CC19E428 CRC64;

Query Match 23.4%; Score 397.5; DB 16; Length 519;
 Best Local Similarity 34.4%; Pred. No. 7.5e-14;
 Matches 109; Conservative 51; Mismatches 128; Indels 29; Gaps 11;

QY 28 SAMVAQVGPQVNNINTKLYNNVAGAGTGIVDPNGWLTNNHVAIGATDINAFSV--GS 85
 Db 205 AGVAAKALPSTVTIQAE-GSNGEGGTGTGFVFDKEGHVITNNHVAEAVDGGKLSATFPN 263
 QY 86 QQTVGVVGVDRDQVAVLQRLGA-GGLPSAAIAGG--VAYGEPVAMGSGGGGTTPR 142
 Db 264 GKKYDAEYVGHQAQGYDAVAVIKLENAPDLKPLALGDSKVAVGDSITAIAGAPFLSNT-- 321
 QY 143 AVPRGVVALGQTVQASDSLGTGAETLNGLIQFDDAIIOPGDSGPPVWVNGLGQVGVNNTA-- 200
 Db 322 VTTGIISAKRPVASSDGSADSKASYMSALQTDASINPNSGGPFLDQAQGVNIGINAIQ 381
 QY 201 --ASDNFQSGGQ---GFAIPICQMAIAGIRSGGSPVTHIGPTAFGLGVVNDNGN 255
 Db 382 STNGGGFTGQAGSTGLGFAIPVNAQFVQAQLKSGKPVVAKIGASVSL-----BETT 436
 QY 256 GARV--QRVVGSG-----APASLGISTGDTVITAVDGAIPNSATAMADALNHHHPGDIS 307
 Db 437 GAKUTEQGVGSDPVEKGPADDAAGLKPQGVITKLDREVDSIDSGTTLIGEITWTHKPGDEV 496
 QY 308 VTWQTKSGGTRTGNVTL 324
 Db 497 VTVE-RGGKQHTAEVTL 512

RESULT 11
 Q8DG87 ID Q8DG87 PRELIMINARY; PRT; 375 AA.
 AC Q8DG87;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Serine protease.
 GN TL2436.
 OS Synecococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22251544; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RL "Complete genome structure of the thermophilic cyanobacterium
 Thermosynechococcus elongatus BP-1";
 DNA Res. 9:123-130(2002).
 DR EMBL; AP005377; BAC0988.1; -.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RA "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RA "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005038; BAC71935.1; -;
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR Protease; Complete proteome.
 KW SEQUENCE 472 AA; 4686 MW; AE04FAA09A3B7E0 CRC64;

Query Match 22.6%; Score 384; DB 16; Length 472;
 Best Local Similarity 32.5%; Pred. No. 3.6e-13;
 Matches 103; Conservative 57; Mismatches 133; Indels 24; Gaps 10;

QY 28 SAMVAQGVQVAVNTKLGYNNAVAGTGTVIDPNCVLTNNHVIAGATDINAFSV--GS 85
 Db 157 AGVAAKALPTVTEIAQSSGEG-GTGTGFVFDKQGHIVTNHVAEAVDGGKLTATFPD 215
 QY 86 GQTYGVVDVGYDRTQDVAVLQLRGA--GGLPSAAIGGG--VAVGEPVVMNGSGGGGTTPR 142
 Db 216 GKYNAEVWGHAGQYDVAVVVKLNAPSDIQLPLGSDKVAVGDSITAIAGAPFLSNT-- 273
 QY 143 AVGRVVALQQTVAQSDSLTCAETLNGLIQFDAAIQPCDSCGVPVNGVGVGNTA-- 200
 Db 274 VTGTGISAKRPVASSDGGSSKASTYMSALQTDASINPGNSGGLDQAQSGVIGNSAIQ 333
 QY 201 ASDNFQLSQGGQ-----GFAIPGQMAIAGQIRSGGSPVTHIGPTAFL-----GLGV 249
 Db 334 SSSSGGLSGSGSGSGLGLFAIPINAKYVAQELIKTGPVPVIGASVLEGTGGAKI 393
 QY 250 VDNNGNGARVQRVVGAPASLGISTGVDITAVDGPAPINSATAMADALNGHHPGVISVT 309
 Db 394 TEQASGGS--DAITPNRPAKAGKPGDVTIKLDDMDVIDSGPTLIGEIMWTRPGATVKLT 451
 QY 310 WQTKSGGTGNTGNVTLAE 326
 Db 452 Y-TRDGKARTDVTLLGE 467

RESULT 14

Q821L8 PRELIMINARY; PRT; 619 AA.
 AC Q821L8;

DT 01-JUN-2003 (TEMBLrel. 24, Created)

DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Putative serine proteinase.
 GN SAV3115.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RA "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RA "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005033; BAC70826.1; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 1.
 KW Complete proteome.
 SQ SEQUENCE 619 AA; 64251 MW; D2D3C72A67D3B3A6 CRC64;

Query Match 22.6%; Score 384; DB 16; Length 619;
 Best Local Similarity 34.0%; Pred. No. 4.9e-13;
 Matches 115; Conservative 53; Mismatches 126; Indels 44; Gaps 12;
 QY 13 SQDRFADFPALPLDPSAMVAQGVQVAVNTKLGYNNAVAGTGTVIDPNCVLTNNHVI 72
 Db 298 SEERAADSV-----GIAARALPSVVTLVHK--GSAAGTGTGFLDGRGHILTNNHV 349
 QY 73 --AGATDINAFSGSGGTGYVDVVDYDRTQDVAVLQLRAGGLPSAAIGG--GVAVGEPV 128
 Db 350 EPAGSGSEISVTFSGGGETAKATVVGGRDSGYDLAVVKVSGVGGKPMPLGNSDNVQGDVP 409
 QY 129 VAMNGSGGGGTTPRAVGRVVALGQTQVA--SDSLTGAETLNGLIQFDAAIQPCDSCGVP 187
 Db 410 VAIGAPDLANT--VTSGLISAKERPTTAGKEKGDGSDVSVVALQTDAPINFGNSGGLP 467
 QY 188 VNGLGQVGVGNTAA-----SDNFQLSQGGQ---GFAIPGQMAIAGQIRSGGSPVTHIG 240
 Db 468 LDSKARVGVGINSAIRSADSSDQSGAGSIGLGLFAIPVNAQKEVABELINTGRATHPVIG 527
 QY 241 PTAFLGLGVVDNNGNGARV-----QRVVGSAPASLGISTGVDITAVDGPAPINSATA 292
 Db 528 VTLDN-----DYTGDGARVGTKNDDGSPVTRGGPDRAIGQAGDVTITEVDGRIHSSEE 582
 QY 293 MADALNGHHPGVDISVTWQTKSGGTRTGN---VTLAEG 327
 Db 583 LIVKIRAHKPDRLALTV-----RDGKEKPVTVLVIG 614

RESULT 15

Q8FRI7

ID	Q8FR17	PRELIMINARY;	PRT;	473 AA.
AC	Q8FR17;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Putative serine protease, heat shock protein.			
GN	CE0950.			
OS	Corynebacterium efficiens.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Corynebacteriaceae; Corynebacterium.			
OX	NCBI_TaxID=152794;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;			
RA	Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,			
RA	Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,			
RA	Usuda Y., Sugimoto S.;			
RT	"The entire genomic sequence of Corynebacterium efficiens YS-314.";			
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AP005217; BAC17760.1; -.			
DR	GO; GO:0003773; F:heat shock protein activity; IEA.			
DR	GO; GO:0008233; F:peptidase activity; IEA.			
DR	GO; GO:004295; F:trypsin activity; IEA.			
DR	GO; GO:007242; P:intracellular signaling cascade; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR009003; Cys_Ser_trypsin.			
DR	InterPro; IPR001478; PDZ.			
DR	InterPro; IPR001254; Peptidase_S1.			
DR	InterPro; IPR001940; Peptidase_S1C.			
DR	Pfam; PF00595; PDZ; 1.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PR00834; PROTEASES2C.			
DR	SMART; SM00228; PDZ; 1.			
DR	PROSITE; PS50106; PDZ; 1.			
KW	Heat shock; Protease; Complete proteome.			
QY	SEQUENCE 473 AA; 47757 MW; 5C0C6C9561F8C524 CRC64;			
Qy	Query Match 22.3%; Score 379; DB 16; Length 473;			
Qy	Best Local Similarity 34.3%; Pred. No. 6.6e-13;			
Qy	Matches 108; Conservative 50; Mismatches 131; Indels 26; Gaps 9;			
Qy	22	ALPLDPSAMVAQYQPVNINTKLYNNVAGTGVDPNGVVLNNHVIAGATDINAF	81	
Db	170	AEFGSAEQVADQVLPSVVSQALT--RNSASEGSGSIISDGFVTNNHVVAGVEQGV	227	
Qy	82	SV--GSGQTYGVVGVYDRTQDVAVLQLRGAGGLPSAAIG--GGVAVGEPVVMGNSGGQ	137	
Db	228	QVTMSDGSSEYADFAGDVSTDIATKIRNATLPMINFGDSLSLVGQEVLA VGSPLGL	287	
Qy	138	GGTPRAVPGRWALGQTVQASDSLTAETINGLIQFDAAIQCDSGGPPVNGLGQVGM	197	
Db	288	SST--VTSGIVSALNRPVRASGD--GGESSLIDAIQTDAAINFGNSGGGLVDMEGNLIGM	343	
Qy	198	NTAASDNFQLSQ-----GQGFAPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVV	250	
Db	344	NSVIA---SISGGTDTGSGIGLGFAPANFAKRVATQLIETGVATQPMIGVTLANGAV-	399	
Qy	251	DNNGNGARQVRVGSAPAAISLGISTGVITAVDGPAINSATAMADALNHHHPGDVISVTW	310	
Db	400	----SGALIASVQGGFGAAGLXPGDVVTMLNDRVIDSADALIAAVRSQDFGETVLTIV	455	
Qy	311	QTK-SGGTRTGNVTL 324		
Db	456	TAQDTSQTRQVEVTL 470		

Search completed: June 22, 2004, 17:22:46
Job time : 27.9826 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 5.49204 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-6
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSQDRPADF.....QTKSGTGTGNVLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	323.5	19.1	458	1 YVTA_BACSU	Q9r9i1 bacillus su
2	323	19.0	413	1 HTRA_LACHE	Q9z4h7 lactobacil
3	316	18.6	355	1 DEGS_ECOLI	P31137 escherichia
4	312	18.4	408	1 HTRA_LACLA	Q91a06 lactococcus
5	305.5	18.0	513	1 DEGP_BRUME	Q8Vg32 bruceella me
6	304.5	17.9	513	1 DEGP_BRUSU	Q44597 bruceella su
7	302.5	17.8	455	1 DEGO_ECOLI	P39099 escherichia
8	293	17.3	475	1 DEGP_SALTY	P26982 salmonella
9	293	17.3	504	1 DEGP_RHIME	Q52894 rhizobium m
10	291.5	17.2	449	1 HTRA_BACSU	Q34358 bacillus su
11	291.5	17.2	474	1 DEGP_ECOLI	P09376 escherichia
12	289	17.0	466	1 HTOA_HAEIN	P45129 haemophilus
13	285.5	16.8	503	1 DEGP_BAHE	P54925 bartonella
14	285	16.8	497	1 DEGP_CHLTR	P18584 chlamydia t
15	284	16.7	478	1 DEGP_BUCAL	P57322 buchnera ap
16	280.5	16.5	497	1 DEGP_CHLMO	Q9p197 chlamydia m
17	279	16.4	488	1 DEGP_CHLPN	Q9z260 chlamydia p
18	278	16.4	437	1 DEGI_ARATH	O22609 arabidopsis
19	277	16.3	508	1 DEGP_RICCN	Q927a1 rickettsia
20	271.5	16.0	400	1 VYXA_BACSU	P39568 bacillus su
21	269.5	15.9	448	1 DEGS_ARATH	Q91u10 arabidopsis
22	267.5	15.8	478	1 DEGP_BUCAP	O65931 buchnera ap
23	251	14.8	513	1 DEGP_RICPR	O05942 rickettsia
24	245.5	14.5	458	1 HRA2_HUMAN	P43464 homo sapien
25	239	14.1	340	1 DEGS_HAEIN	Q92743 homo sapien
26	232.5	13.7	480	1 HRA1_HUMAN	P83110 homo sapien
27	231.5	13.6	453	1 HRA3_HUMAN	Q9j1v5 mus sapien
28	225.5	13.3	458	1 HRA2_MOUSE	P83105 mus sapien
29	222	13.1	476	1 HRA4_HUMAN	Q9d336 mus sapien
30	218.5	12.9	460	1 HRA3_MOUSE	Q9r1i8 mus sapien
31	218	12.8	480	1 HRA1_MOUSE	Q9se17 arabidopsis
32	175.5	10.3	321	1 SPPA_ARATH	P55377 rhizobium s
33	172.5	10.2	630	1 Y4BJ_RHISN	

RESULT 1

ID	YVTA_BACSU	STANDARD;	PRT;	458 AA.
AC	Q9r9i1; Q35021; Q35039;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Probable serine protease yvta (EC 3.4.21.-).			
GN	YVTA OR YVTB OR BSU33000.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=20158875; PubMed=10692364;			
RA	Noone D., Howell A., Devine K.M.;			
RT	"Expression of ykda, encoding a Bacillus subtilis homologue of Htra,			
RT	is heat shock inducible and negatively autoregulated.";			
RL	J. Bacteriol. 182:1592-1599(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98015415; PubMed=9353931;			
RA	Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;			
RT	"Sequencing of regions downstream of addA (98 degrees) and citG (289			
RT	degrees) in Bacillus subtilis.";			
RL	Microbiology 143:3305-3308(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulic M.P., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Puig J.C., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solido B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tamakoehi A., Taraka T., Terpsstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			

Q64761 avian adeno
P04985 bos taurus
Q50594 mycobacteri
Q99372 rattus norv
P52320 streptomyc
P56877 mycobacteri
Q53810 mycobacteri
O06794 mycobacteri
Q52657 rickettsia
P12031 sus scrofa
Q53047 r outer mem
O06653 r outer mem

ALIGNMENTS

[illegible]

RESULT 3

DEGS_ECOLI	DEGS_ECOLI	STANDARD;	PRT;	355 AA.
ID	AC	P31137;		
DT	01-JUL-1993	(Rel. 26, Created)		
DT	01-FEB-1995	(Rel. 31, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Protease degs precursor	(EC 3.4.21.-)		
GN	DEGS OR HHOB OR HPRH OR B3235 OR Z4594 OR ECS4108.			
OS	Escherichia coli, and			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562, 83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / W3110;			
RA	Bass S., Gu Q., Goddard A.;			
RA	Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / W3110;			
RC	MEDLINE=96165272; PubMed=8576051;			
EX	Waller P.R., Sauer R.T.;			
RT	"Characterization of degQ and degS, Escherichia coli genes encoding			
RT	homologs of the DegP protease."			
RL	J. Bacteriol. 178:1146-1153(1996).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RC	MEDLINE=97426617; PubMed=9278503;			
EX	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregg J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12."			
RL	Science 277:1453-1474(1997).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RC	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregg J., Kirkpatrick H.A.,			
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,			
RA	Apodaca J., Arantharuman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Welch R.A., Blattner F.R.;			
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."			
RL	Nature 409:529-533(2001).			
RN	[5]			

[illegible]

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AE009571; AAL52511.1; --

PIR; AD3418; AD3418; Cys_Ser_trypsin.

InterPro; IPR003003; Cys_Ser_trypsin.

InterPro; IPR001478; PDZ.

InterPro; IPR001254; Peptidase_S1.

InterPro; IPR001940; Peptidase_S1C.

Pfam; PF00595; PDZ; 2.

Pfam; PF00089; trypsin; 1.

PRINTS; PR00834; PROTEASES2C.

SMART; SM00228; PDZ; 2.

PROSITE; PS0106; PDZ; 2.

Hydrolase; Serine protease; Periplasmic; Repeat; Signal; Complete proteome.

KW SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 513 PROBABLE SERINE PROTEASE DO-LIKE.

FT DOMAIN 125 299 CATALYTIC.

FT DOMAIN 300 391 PDZ 1.

FT DOMAIN 414 500 PDZ 2.

FT ACT_SITE 152 152 CHARGE RELAY SYSTEM (POTENTIAL).

FT ACT_SITE 182 182 CHARGE RELAY SYSTEM (POTENTIAL).

FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (POTENTIAL).

SQ SEQUENCE 513 AA; 53514 MW; C472FEF9DFC6268 CRC64;

Query Match 18.0%; Score 305.5; DB 1; Length 513;

Best Local Similarity 34.0%; Pred. No. 1e-10;

Matches 103; Conservative 37; Mismatches 118; Indels 45; Gaps 13;

QY 46 GYNNAVGAGTGVIDPNGVLTNNHVNAGATDINAFSV--GSGQTYGVVDVGYDRTQDVA 103

DB 128 GHERPVAQSGGFVISEDGVYVNNHVV--SDGDAYTVLLDDGTDLAKLIGADPRTDLA 184

QY 104 VLQIRAGGHLPS-----AATG--GGVAVGEPVVMANGSGGGCTPPRAVPGHVWALGQTQV 156

DB 185 VLKINA----PKRFVTVAFGDDNKNRVGDMVAVGNPFGLGGT--VTSGLIVSARGRDIG 238

QY 157 ASDSLTCAEETLNLGIQDAIOPGDAGGPPVNGLGQVGVGMNTAASDNFQLSGQGQ--F 214

DB 239 AG-----PYDDFIQIDAVNKGNSGGFADLSGEVIGINTAI---FSPSGSGVGIAP 287

QY 215 AIPFIGQAWIAGQIRSGG----GSPTVHIGP-----TAFGLGVVDNNGNARGVORVVGVS 265

DB 288 AIPSTAKQVVDQLIKKGSVERGWIGVQIQPVTKDIAASGLA---EEKGAIVASPQDD 343

QY 266 APAASLGISGTDVITAYDGPINISATAMADALNGHPCDVISVTWQKSGGTRIGNVTLA 325

DB 344 GPAAKAGIKAGDVITAVNGETVQDPRLARKVANIAPGEKAALTVRNKKAEEI-NVTIIA 402

QY 326 EGP 328

DB 403 AMP 405

RESULT 6

DEGP BRUSU STANDARD; PRT; 513 AA.

AC Q44597;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable serine protease do-like precursor (EC 3.4.21.-).

GN DEGP OR HTRA OR BR0611.

CS Brucella suis, and

OS Brucella abortus.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Brucellaceae; Brucella.

OX NCBI_TaxID=29461, 235;

RN [1]

RN SEQUENCE FROM N.A.

RP SPECIES=B.abortus; STRAIN=2308;

RC

Db 288 AIPSTAKQVVDLIKKSGVERGWIGVQIQVTKDKIAASLGLA-----EKKGAIVASPDQD 343
 QY 266 APAASLGISTGVDITAVDGAIPINSATAMADALNGHHPGDVVISVTWTKSGGRTGNVTIA 325
 Db 344 GPAKAGIKAGDITAVNGETVQDPRDLAKKVNIAPEKAALTVWRKKABEI-NVTIA 402
 QY 326 EGP 328
 Db 403 AMP 405

RESULT 7
 DEQG_ECOLI STANDARD; PRT; 455 AA.
 AC P39099;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE "Protease degQ precursor (EC 3.4.21.-).
 GN DEQG OR HHOA OR B3234.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Bass S., Gu Q., Goddard A.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RT "Characterization of degQ and degS, Escherichia coli genes encoding
 RT homologs of the DegP protease."
 J. Bacteriol. 178:1148-1153(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MGL655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 Science 277:1453-1474(1997).
 RL CC -1- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGQ/DEGP.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: Belongs to peptidase family S2C.
 CC -1- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U15661; AAC3992.1; -;
 DR EMBL; U32495; AAC44005.1; -;
 DR EMBL; U18997; AAA58036.1; -;
 DR EMBL; AE00402; AAC76266.1; -;
 DR PIR; JC6051; JC6051.
 DR MEROPS; S01.274; -;
 DR SWISS-2DPAGE; P39099; COLI.
 DR EcoGene; EG12612; degQ.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.

DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 KW Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 455
 FT DOMAIN 258 349
 FT DOMAIN 355 447
 FT ACT_SITE 109 109
 FT ACT_SITE 139 139
 FT ACT_SITE 214 214
 FT ACT_SITE 214 214
 SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;
 Query Match 17.8%; Score 302.5; DB 1; Length 455;
 Best Local Similarity 29.4%; Pred. No. 1.4e-10;
 Matches 100; Conservative 50; Mismatches 101; Indels 89; Gaps 14;
 QY 15 DRFADFFALPLDPSAMVAQVGPQVNVINTKLGYNVAGAGTGVIDEN-GVLTNNHIVIA 73
 Db 78 DDLDPQAPQPE-----GLGSGVIINASKGVYTNHVIN 112
 QY 74 GATDINAFSGSGGTGVVGVYDRTQDVAVLQRLGAGGLPSAAIGGG--VAVGEPVVM 131
 Db 113 QAQKI-SIQLNDGREFDAKLIGSDQSDIALQLQIONPSKLTQIAIADSKLRVGDFAVAV 171
 QY 132 GNSGGGGTTPRAVGRVVALGVQVQASDSLTGAB-ETLNGLIQPDAAIQPDAGGPVNG 190
 Db 172 GNPFLGQT--ATSGIVSALGR-----SGLNLEGLENFQTDASINRGSGGALLNL 221
 QY 191 LGQVVGWNTAASDNFOLSQGG--GFAIPICGMAIA-----GQIRSGSGSPVTHIG 240
 Db 222 NGELIGNTAI-----LAPGGSGVIGFALPNSNARTLAQQLIDFGKIKG----- 267
 QY 241 PTAFLGLGVVDNNGN-----GARVQRVVGSAAPASLGISTGDTAVDGAIPINS 289
 Db 268 ---LLGIGKTEMSADIAKAFNLDVQRCGAFVSEVLPGSGSAKAGVKAGDIITSLNGKPLNS 324
 QY 290 ATAMADALNGHHPGDVSVTWTKSGGTRTG-----NVTIL 324
 Db 325 FAELRSRIATTEPGT-----KVKLGLLRNGKPLEVEVTL 358

RESULT 8
 DEGP_SALTY STANDARD; PRT; 475 AA.
 AC P26982;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protease do precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR PTD OR STM0209.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C5;
 RX MEDLINE=91251770; PubMed=1645840;
 RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
 RA Ali T., Miller I., Hormaeche C.;
 RT "The role of a stress-response protein in Salmonella typhimurium
 RL virulence."
 RL Mol. Microbiol. 5:401-407(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 67 POTENTIAL.
FT DOMAIN 68 449 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 103 108 POLY-SER.
FT DOMAIN 146 152 POLY-SER.
FT DOMAIN 348 437 POZ.
FT ACT_SITE 179 179 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 209 209 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 449 AA; 47713 MW; E12B07A9018EB414 CRC64;
Query Match 17.2%; Score 291.5; DB 1; Length 449;
Best Local Similarity 29.0%; Pred.No.5.7e-10;
Matches 92; Conservative 53; Mismatches 127; Indels 45; Gaps 11;
QY 28 SAMVAQVGPQVNNITKLYGNNNA-----VAGAGTGTVI-DPENG--VVLINN 69
Db 119 SDWVEDLSPAIVGITLQAGNSLSFGSSSDSDSTSGSGGVIFKXGKNGKAVIINN 178
QY 70 HVIAGATDINARVSGSGQYGVVDVYQDRTQVAVLQLRAGAGLPSAIG--GGVAVGEP 127
Db 179 HVVEGASSLKV-SLYDGTGETVAKLVGSDSLTLAVLQISDDHVTYKVFNGSDSLRTGET 237
QY 128 VVAMGNSGGQGTTPRAVPGRVVAGVQVQASDLSLTGAERTNLGLQFDAAIQFDAGGVP 187
Db 238 VIAIGDPLGKDLRSRTVQGVISGVDTVSM-----TSAGETSNVITQDAINFGSGGPL 294
QY 188 VNLGQVGVGNT--AASDNFQISQGGQPAIPIGQMAIAGQIRSGGSGPTVHIGTAPL 245
Db 295 LNTDGRIVGINSMKISDDVE-----GIGFAIPSDNDVKPIAEELSKGQTEREYIG-VSML 349
QY 246 GLGVDDNN-----NGARVQVWGSAPASIGTGVITAVDAPINSATA 292
Db 350 DLEQVQNTQEGTLGLFGSLKNGVYIREVAGSGFAEKAGLKAEDIIIGLKREIDTUSE 409
QY 293 MADAL-NGHHPGDVTSV 308
Db 410 LRNLYKDAKIGDTVEV 426
RESULT 11
DEGP_ECOLI STANDARD; PRT; 474 AA.
AC P09376; P15724;
DT 01-MAR-1989 (Rel. 10, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.
OS Escherichia coli, and
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCB_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htra gene of Escherichia coli: a sigma 32-independent mechanism of heat-inducible transcription."
RL Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Kura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the 2.4-4.1 min (110,917-193,643 bp) region."
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E., Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D., Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobick E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7." Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12." DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90323597; PubMed=2165018;
RA Quirk S., Bhatnagar S.K., Bessan M.J.;
RT "Primary structure of the deoxyguanosine triphosphate triphosphohydrolase-encoding gene (dgt) of Escherichia coli." Gene 89:13-18(1990).
RN [8]
RP SEQUENCE OF 1-16 FROM N.A.
RX MEDLINE=90207273; PubMed=2157212;
RA Wurgler S.M., Richardson C.C.;
RT "Structure and regulation of the gene for dGTP triphosphohydrolase from Escherichia coli." Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
RN [9]
RP SEQUENCE OF 27-39, AND CHARACTERIZATION.
RX MEDLINE=90202693; PubMed=2180903;
RA Lipinska B., Zylicz M., Georgopoulos C.;
RT "The htra (DegP) protein, essential for Escherichia coli survival at high temperatures, is an endopeptidase." J. Bacteriol. 172:1791-1797(1990).
RN [10]
RP SEQUENCE OF 27-30.
RC STRAIN=K12 / W3110;
RX MEDLINE=98263247; PubMed=9600841;
RA Wilkins M.R., Gasteiger E., Tonella L., Ou K., Tyler M., Sanchez J.-C., Goolley A.A., Walsh B.J., Bairoch A., Appel R.D., Williams K.L., Hochstrasser D.F.;
RT "Protein identification with N and C-terminal sequence tags in proteome projects." J. Mol. Biol. 278:599-608(1998).
RN [11]
RP IDENTITY OF HTRA AND PROTEASE DO.
RX MEDLINE=91222240; PubMed=2025286;

SEQ	SEQUENCE	474 AA; 49354 MW; 5482E596F74B6D5F CRC64;	
	Query Match	17.2%; Score 291.5; DB 1; Length 474;	
	Best Local Similarity	33.2%; Pred. No. 6e-10;	
	Matches	88; Conservative 43; Mismatches 81; Indels 53; Gaps 11;	
QY	54	GTGIVIDPN-GVLTNNHVIAGATDINAFSVSGSGQTYGVDVGYDRQTQDAVQLRGAAG 112	
DB	114	GSQVIIDADGYVTNNHVVDNATVIKV-QLSGRKFDKMGVDPKPSDIALIQTQNPKN 172	
QY	113	LPS--AAIGGGVAVGPEVVMGNSGGQGTTPRAVPGRVVALGQTVQASDSLTCAG-ETLN 169	
DB	173	LTAIKWADSDALRVGDTVAIGNPFLGET--VTSGIVSALGR-----SGLNAENYE 222	
QY	170	GLIQPDAAITQPDAGGPEVNGLVGVGMNTA--ASDNFOLSGQGQGAIP-----IGQA 221	
DB	223	NTIQTDAALNRGNSGALVNLGELIGINTAILAPDGGNI--GIGFALPSNMVKNLTQS 279	
QY	222	MAIAGQIRSGGSGPTVHIQPTAFLGLVGVNNGN-----GARQVQVGSAPA 268	
DB	280	MVEYGVQVKG-----ELGIMGTEINSELAKMKVDAQRGAFVSQVLPNSSA 325	
QY	269	ASIGISTGVIVITAVDGAIPNSATAM 293	
DB	326	AKAGIKAGDVITSLNGKPISSPAAL 350	
RESULT 12			
HTQA_HA3IN			
ID	HTQA_HA3IN	STANDARD; PAT; 466 AA.	
AC	P45129;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DE	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Probable periplasmic serine protease do/hhoA-like precursor		
DE	(EC 3.4.21.-).		
GN	Hli259.		
OS	Haemophilus influenzae.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;		
OC	Pasteurellaceae; Haemophilus.		
ON	NCBI_TaxID=727;		
OR	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Rd / KW20 / ATCC 51907;		
RC	MEDLINE=95350630; PubMed=7542800;		
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,		
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,		
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,		
RA	Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,		
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,		
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,		
RA	Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,		
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,		
RA	Venter J.C.;		
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae		
RT	Rd,"		
RL	Science 269:496-512(1995).		
CC	-1- SUBCELLULAR LOCATION: Periplasmic (Potential).		
CC	-1- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA		
CC	(PROTEASE DO) AND HHOA.		
CC	-1- SIMILARITY: Belongs to peptidase family S2C.		
CC	-1- SIMILARITY: Contains 2 PDZ/DHR domains.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/)		
CC	or send an email to license@isb-sib.ch .		
CC			
CC	EMBL; U32805; AAC22906.1; -.		
DR	PIR; A64113; A64113.		

```

DR MEROPS; S01.274; -.
DR TIGR; H11259; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR HydroLase; Serine protease; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 466 DO/HOAA-LIKE.
FT DOMAIN 270 361 PDZ 1.
FT DOMAIN 367 458 PDZ 2.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 466 AA; 49433 MW; ED050A00047B5851 CRC64;

Query Match 17.0%; Score 289; DB 1; Length 466;
Best Local Similarity 32.6%; Pred. No. 8.2e-10;
Matches 93; Conservative 43; Mismatches 101; Indels 48; Gaps 11;

QY 49 NAVGAGTGIVDPN-QVLTNNHVIAGTDINAFVSGSQTYGVVDVDRDQDVAVLQL 107
DB 98 NFRGLSGVIIINASKGYVLTNNHVIDGADKITV-QLQDGRFKAKLVGKDEQSDIALVQL 156
QY 108 RGAGGLPSA--AIGGGVAVGEPVWAMGNSGGGGTTPRAVPGRVWALGQTVQASDSLTGAE 165
DB 157 EKPNSLTKFKPADSKLRVGDFTVAIGNFGLGQT--VTSGIVSALGRS-----TGSD 207
QY 166 E-TLNGLIQDAAIQDAGGVPVNGLVQVGMNTAASDNFQLSGGGG--FAPIGQAM 222
DB 208 SGTENYIQTDAAVNRGSGGALVNLGELIGINTAI---ISPSGNAGIAFAIPSNQAS 264
QY 223 ATA-----GQIRSGGGSVTIHGPTAFLGLGVNDNGN-----GARVQVVS 265
DB 265 NLVQQLIEFGVRRG-----LLGKGGELNADLAKAFNVSAQQGAFVSEVLPK 312
QY 266 APAASLGISTGVITAVDGAPINSATAMADALNGHHFGSDVISVTW 310
DB 313 SAAEXAGLRAGDIITAMNQSQKISSFAIRAKIATTGAGKEISLTY 357

RESULT 13
DEGP_BARHE STANDARD; PRT; 503 AA.
AC P54925;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
DE (Antigen H7A).
GN DEGP OR H7A.
OS Bartonella henselae (Rochalimaea henselae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=38323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49882 / Houston 1;
RX MEDLINE=94299828; PubMed=8027347;
RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
RA Goral S., Hager C., Edwards K.;
RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch
RT disease patients by PCR."
RL J. Clin. Microbiol. 32:942-948(1994).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.

```

```

CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L20127; AAA97430.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW HydroLase; Serine protease; Periplasmic; Repeat; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 503 PROBABLE PERIPLASMIC SERINE PROTEASE DO-
FT DOMAIN 286 357 PDZ 1.
FT DOMAIN 419 466 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 503 AA; 54114 MW; 6CD9F4743282AF9E CRC64;

Query Match 16.8%; Score 285.5; DB 1; Length 503;
Best Local Similarity 27.4%; Pred. No. 1.4e-09;
Matches 107; Conservative 43; Mismatches 125; Indels 115; Gaps 15;

QY 12 LSQDRFADFALPLDPSAMVAQVGPVQVNTKLG----- 46
DB 47 MQQQGFAD-----IVSQVKPAVVSVQVSKNKKKEWFFSDFTSTPGFDQLPDQHP 96
QY 47 -----YN-----NAVAGTGVIVDPNGVLTNNHVIAGTDINAFSV 83
DB 97 LKFFQDFYNRDKPSNKLQRSHLRPIAFSGSFSSDGYIVTNNHVISDGTSY-AVVL 155
QY 84 GSGQTYGVVDVVDYDRTQDVAVLQLRGAGGLPSAAIG--GGVAVGEPVWAMGNSGGGGTTP 141
DB 156 DGTGLNAKLIGTPTDRLAVLKVNEKFSYVDYFGDDSKLRVGDVWVAIGNPFLGTV 215
QY 142 RAVPGRVVVALGQTVQASDSLTGAEETLNGLIQDAAIQDAGGVPVNGLVQVGMNTAA 201
DB 216 TA--GIVSARG-----DIGTG---VYDDFIQIDAAVNRGNSGGFTFDLNGKVYGVNTAI 265
QY 202 SDNFQLSGGGG--FAPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARV 259
DB 266 ---FSPGGNVGIAFAIPAATANEVQQL-----IEKGLVQRWLGVOI 306
QY 260 QRV-----VG-----SAPASLSGISTGDEVITAVDGAFINSATAMADALNG 299
DB 307 QPVTKEITSDSIGLKEAKGALITDPLKGPAAKAGIKAGDVIISVNGEKINDVRDLAKRIAN 366
QY 300 HPGDVISV-TWQTKSGGTRTGNVTLAEGP 328
DB 367 MSPGETVLGW--KSGKEENIKVKLDSMP 394

RESULT 14
DEGP_CHLTR STANDARD; PRT; 497 AA.
AC P18584; O84830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa

```

	QY	114	P\$AAITGG--VAVGPPVVAMGNKSGCGGTTPRAVPGEVALQT-----VOASDLTGAEET	167
	DB	185	PFLTFGNSDQLIGDAIWAIAIGNPFLQAAT-VTVGVISAKGRNQLHIVDFED-----	234
	QY	168	LNLGIQFDAAIQPGDAGGPVNVLGQVGVMNTAASNFLQSOG--CQGFAIFIQQAMAIA	225
	DB	235	---FIOTDAAINPNNGSGGPELLNINGQVICWTAIVSG---SCGYIGIFAIPLMAKRVI	288
	QY	226	GQIRSGGSPTWHIGTAPLGLGVVDN-----NNGARVQRVVGSPAASLGISTG	276
	DB	289	DQLSIDSQVTGRFGLVLT----LPIDSELAATCYKEKVYGALVTDVKSGSPAKAGLRQE	344
	QY	277	DVITAADGPAINSAYAMADALNGHHHPGDVISWTQTSKGTRTNVTLAEGP	328
	DB	345	DIVVAENGKEVESLSALRNAISLMMPGTRV-ULKIVRECKTIIEIPVTVQIP	395
			RESULT 15	
	ID	DEGP_BUCAI	STANDARD; PRT; 478 AA.	
	AC	P5732;		
	DT	16-OCT-2001 (Rel. 40, Created)		
	DT	16-OCT-2001 (Rel. 40, Last sequence update)		
	DT	10-OCT-2003 (Rel. 42, Last annotation update)		
	DN	Probable serine protease do-like precursor (EC 3.4.21.-).		
	GE	DEGP OR BU228.		
	OS	Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).		
	OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
	OC	Enterobacteriaceae; Buchnera.		
	OX	NCBI_taxid=118099;		
	[1]			
	RN	SEQUENCE FROM N.A.		
	RP	STRAIN=Tokyo 1998;		
	RC	MEDLINE=20445173; PubMed=10993077;		
	RX	Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;		
	RA	"Genome sequence of the endocellular bacterial symbiont of aphids		
	RT	Buchnera sp. APS.";		
	RL	Nature 407:81-86(2000).		
	-!	SIMILARITY: Belongs to peptidase family S2C.		
	--!	SIMILARITY: Contains 2 PDZ/DHR domains.		
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).		
	CC	EMBL; AP001118; BAB12943.1; --		
	DR	MEROPS; S01.273; --		
	DR	InterPro; IPR009003; Cys_Ser_trypsin.		
	DR	InterPro; IPR001478; PDZ.		
	DR	InterPro; IPR001254; Peptidase_S1.		
	DR	InterPro; IPR001940; Peptidase_S1c.		
	DR	pfam; PF00595; PDZ; 2.		
	DR	pfam; PF00089; trypsin; 1.		
	DR	PRINTS; PRO0834; PROTEASES2C.		
	DR	SMART; SM00228; PDZ; 2.		
	DR	PROSITE; PS50106; PDZ; 1.		
	KW	Hydrolase; Serine protease; Repeat; Signal; Complete proteome.		
	F	Hydrolase; Serine protease; Repeat; Signal; Complete proteome.		
	T	CHAIN 1 26 POTENTIAL,		
	FT	CHAIN 27 478 PROBABLE SERINE PROTEASE DO-LIKE.		
	FT	DOMAIN 116 254 CATALYTIC.		
	FT	DOMAIN 281 372 PDZ 1.		
	FT	DOMAIN 387 469 PDZ 2.		
	FT	ACT_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).		
	FT	ACT_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).		
	FT	ACT_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).		
	FT	DISULFID 57 99 BY SIMILARITY.		
	SEQ	SEQUENCE 478 AA; 52230 MW; 869E8732CAC50629 CRC64;		

Query Match 16.7%; Score 284; DB 1; Length 478;
Best Local Similarity 30.6%; Pred. No. 1.6e-09;
Matches 88; Conservative 51; Mismatches 109; Indels 40; Gaps 12;

QY	54	GTGIVIDPN-GVVLNNHVIAGATDINAFSVGSGOTYGVVDVGYDRTOADVAVLQLRGAGG	112
Db	116	GGVVIINADKGYAVTNHHVVENANKIQV-QLSGRRYEAVVIGKDSRDIALIQLKNANN	174
QY	113	LPSAAI--GGGVAVGEFVVAMGNSGGQGTTPRAVGRVVVALGQTVQASDSLGTAE-ETLN	169
Db	175	LSEIKIADSDNLRVGDYTVAGNPYGLGET--VTSGIISALGR-----SGLNIEHYE	224
QY	170	GLIQDAAIOPGDAGPVGNGVGVGMNTA--ASDNFQLSQGGGPAIPGQAMAIAGQ	227
Db	225	NFIQDAAINRNSGGALVNLKGLIGINTALADGGNI---GIGFALPCNMVKNLTAQ	281
QY	228	IRSGGGSPTVHIGPTAFLGLGVVD-----NNGNGARVQVRVGSAPAASLGISTGDVI	279
Db	282	MYQFG--QVRRGELGIMGELNSDLAQIMKINSQKGFVSRVLPNSSAFAGIKAGDII	338
QY	280	TAVDGAIPNSATAMADALNGHHPGDVISVTWQTKSGGTRTG---NVTL	324
Db	339	ISLNRKPISFSLSRAEIGS-----LPVATKVELGVFREGRIKNITY	380

Search completed: June 22, 2004, 17:16:48
Job time : 6.58295 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 26.9826 Seconds
(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349A-6

Perfect score: 1698

Sequence: 1 MHVHHHAPPALSQDRFADF.....QTKSGGTRTGNVTLAEGPPA 330

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1642	96.7	355	16 O07175	O07175 mycobacteri
2	1642	96.7	355	16 Q7U2S9	Q7U2S9 mycobacteri
3	1228.5	72.3	361	2 Q50320	Q50320 mycobacteri
4	1195	70.4	354	16 Q9CCY9	Q9CCY9 mycobacteri
5	457.5	26.9	446	16 Q8VKA4	Q8VKA4 mycobacteri
6	457.5	26.9	464	16 Q7U0X2	Q7U0X2 mycobacteri
7	456.5	26.9	464	16 Q53896	Q53896 mycobacteri
8	447	26.3	382	16 Q9CD67	Q9CD67 mycobacteri
9	447	26.3	452	2 Q925G6	Q925G6 mycobacteri
10	394.5	23.2	519	16 Q93J30	Q93J30 streptomyc
11	393	23.1	375	16 Q8DC87	Q8DC87 synchococc
12	390.5	22.4	542	16 Q9FBK9	Q9FBK9 streptomyc
13	381	22.0	472	16 Q82FM9	Q82FM9 streptomyc
14	381	22.4	619	16 Q82IL8	Q82IL8 streptomyc
15	376	22.1	473	16 Q8FR17	Q8FR17 corynebacte
16	371	21.8	394	16 P72780	P72780 synchocyst

ALIGNMENTS

RESULT 1

ID	007175	PRELIMINARY;	PRT;	355 AA.
AC	O07175;			
DT	01-JUL-1997 (Tremblrel. 04, Created)			
DT	01-JUL-1997 (Tremblrel. 04, Last sequence update)			
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)			
DE	Hypothetical protein (Serine protease, putative).			
GN	PEPA OR RV0125 OR MTC1418B.07 OR MT0133.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Decher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			

Q8ytf9 anabaena sp
Q8dmv9 synchococc
Q8gat3 bifidobacte
Q8na10 corynebacte
Q7vsc8 prochloroco
Q89i15 clostridium
Q97gd5 clostridium
Q8r756 thermoanaer
P73354 synchocyst
Q89rp2 bradyrhizob
Q8pmv4 xanthomonas
Q9kjn6 myxococcus
Q7u495 synchococc
Q8xpt5 xalstonia s
Q8dl28 synchococc
Q31388 bradyrhizob
Q89j18 bradyrhizob
Q7u8k9 synchococc
Q87c10 xylella fas
Q9ba3 xylella fas
Q7u604 synchococc
Q9szr5 streptomyc
Q92qe6 rhizobium m
Q8rtk2 xanthomonas
Q8pb56 xanthomonas
Q8yvh0 anabaena sp
Q83752 treponema p
Q9rtk4 deiococcus
Q7va24 prochloroco

CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; Z96071; CAB09453.1; -.
 DR RA "The complete genome sequence of Mycobacterium bovis."
 DR RT "The complete genome sequence of Mycobacterium bovis."
 DR RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR DR EMBL; BX248334; CAD92991.1; -.
 DR KW Hydrolyase; Complete proteome.
 DR SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 96.7%; Score 1642; DB 16; Length 355;
 Best Local Similarity 99.7%; Pred. No. 7.4e-81;
 Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGVIDPENGVL 67
 DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGVIDPENGVL 92
 QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAGGGVAVGEP 127
 DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAGGGVAVGEP 152
 QY 128 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAIIPGDAGGPV 187
 DB 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAIIPGDAGGPV 212
 QY 188 VVNGIQQVVMNTAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGL 247
 DB 213 VVNGIQQVVMNTAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGL 272
 QY 248 GVVDNNGGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
 DB 273 GVVDNNGGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355

Query Match 96.7%; Score 1642; DB 16; Length 355;
 Best Local Similarity 99.7%; Pred. No. 7.4e-81;
 Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGVIDPENGVL 67
 DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGVIDPENGVL 92
 QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAGGGVAVGEP 127
 DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAGGGVAVGEP 152
 QY 128 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAIIPGDAGGPV 187
 DB 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAIIPGDAGGPV 212
 QY 188 VVNGIQQVVMNTAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGL 247
 DB 213 VVNGIQQVVMNTAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGL 272
 QY 248 GVVDNNGGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
 DB 273 GVVDNNGGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2
 Q7U2S9 PRELIMINARY; PRT; 355 AA.
 AC Q7U2S9
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Probable serine protease PEPA (EC 3.4.21.-).
 GN PEPA OR MB0130.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garner T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor W., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,

RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis."
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 RL EMBL; BX248334; CAD92991.1; -.
 DR KW Hydrolyase; Complete proteome.
 DR SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 96.7%; Score 1642; DB 16; Length 355;
 Best Local Similarity 99.7%; Pred. No. 7.4e-81;
 Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGVIDPENGVL 67
 DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGVIDPENGVL 92
 QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAGGGVAVGEP 127
 DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAGGGVAVGEP 152
 QY 128 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAIIPGDAGGPV 187
 DB 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAIIPGDAGGPV 212
 QY 188 VVNGIQQVVMNTAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGL 247
 DB 213 VVNGIQQVVMNTAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGL 272
 QY 248 GVVDNNGGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
 DB 273 GVVDNNGGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3
 Q50320 PRELIMINARY; PRT; 361 AA.
 AC Q50320
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE 34KDa protein precursor.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1770;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=JD88/107;
 RX MEDLINE=95005449; PubMed=7921248;
 RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
 RT "Identification and characterisation of a putative serine protease
 expressed in vivo by Mycobacterium avium subsp paratuberculosis."
 RL Microbiology 140:1977-1982(1994).
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 DR EMBL; Z23092; CAB80638.1; -.
 DR FIR; S47170; S47170.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IP0009003; Cys_Ser_trypsin.
 DR InterPro; IP001478; PDZ_S1.
 DR InterPro; IP001254; Peptidase_S1.
 DR InterPro; IP001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.

KW	Hydrolase; Protease; Serine protease; Signal.					
FT	SIGNAL	1	38	POTENTIAL.		
SQ	SEQUENCE	361 AA;	35709 MW;	30FEF7DFD6F3C411 CRC64;		
	Query Watch	72.3%;	Score 1228.5;	DB 2;	Length 361;	
	Best Local Similarity	72.8%;	Pred. No. 1.3e-58;			
	Matches 235;	Conservative 39;	Mismatches 48;	Indels 1;	Gaps 1;	
QY	8 APPALSDQRADFAPLPLDPSSAMVAQGPQVAINTKLGYNNAVAGAGTGIVIDPNGVLT	67				
Db	40 AFSGLADRFADRPLAPIDPSAMVGQVPQVNIIDTKFGYNNAVAGAGTGIVIDPNGVLT	99				
QY	68 NNHVITAGATDINARVSGGQTGYDVUVGYDRTDQDAVLQLRGAGGLPSAIIGGGVAVGE	127				
Db	100 NNHVISGATEISAFEDVGNQGTYAVDVVVGYDRTDQDAVLQLRGAGGLPTATIGGEATVGE	159				
QY	128 VVMANGSQQGTTPRAVEGRVVALGQTVAQSDSLTAETTLNGLIQFDAAIQPGDAGGPV	187				
Db	160 IVALGNVGGGTPNAVAGKVVALNQSVSATDILTGAENIAGLIQADAPIKPGDSGGPM	219				
QY	188 NVNLGQVGMNTAAASDNFQLSGQGFAIPDGMAIAGQIRSGGSPVTHIGPTAFILGL	247				
Db	220 VNSAQGVTVDTAATDSYFKM-GGQGFAPDIGRAMAVANQIRSGAGSNVTHTGPTAFILGL	278				
QY	248 GYVDNNNGARVQRVGSAPAASLCIGTGVITAVDGAPINSATAMADALNHGHFGDDVIS	307				
Db	279 GYVDNNNGARVQRVVNTGPAAGAIPGDVITGVTVTPINGATSMTVELVPHFGDIIA	338				
QY	308 VTWQTKSGGTTRGVNTLAEGPPA 330					
Db	339 VHFRSVDGERTANTITLAEGPPA 361					
 RESULT 4 Q9CCY9 PRELIMINARY; PRT; 354 AA. 						
ID	Q9CCY9	AC	Q9CCY9	AC	Q9CCY9	AC
DT	01-JUN-2001 (TrEMBLrel. 17, Created)					
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)					
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)					
DE	Probable secreted serine protease.					
GN	ML2659.					
OS	Mycobacterium leprae.					
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;					
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.					
OX	NCBI_TaxID=1769;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=TN;					
RC	MEDLINE=21128732; PubMed=11234002;					
RA	Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,					
RA	Wheeler P.R., Honore N.K., Garnier T., Churcher C., Harris D.,					
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,					
RA	Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,					
RA	Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,					
RA	Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,					
RA	Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,					
RA	Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,					
RA	Barrell B.G.;					
RL	"Massive gene decay in the leprosy bacillus.";					
RT	Nature 409:1007-1011(2001)					
CC	-!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.					
DR	EMBL; ALU583926; CAC32191.1; -.					
DR	FIR; A87242; A87242.					
DR	Leproma; ML2659; -.					
DR	GO; GO:0004295; F:trypsin activity; IEA.					
DR	GO; GO:0007242; P:intracellular signaling cascade; IEA.					
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.					
DR	InterPro; IPRO09003; Cys_Ser_trypsin.					
DR	InterPro; IPRO01478; PDZ.					
DR	InterPro; IPRO01254; Peptidase S1.					
DR	InterPro; IPRO08256; Peptidase S1B v8.					
DR	InterPro; IPRO01940; Peptidase S1C.					

DR	Pfam; PF00595; PDZ; 1.
DR	Pfam; PF0089; trypsin; 1.
DR	PRINTS; PRO0834; PROTEASES2C.
DR	PRINTS; PRO0839; V8PROTEASE.
DR	SMART; SMC0228; PDZ; 1.
DR	PROSITE; PS50106; PDZ; 1.
KW	Hydrolase; Protease; Serine protease; Complete proteome.
SQ	SEQUENCE 354 AA; 32665 MW; 612F23261BC9EA4A CRC64;
	Query Match 70.4%; Score 1195; DB 16; Length 354;
	Best Local Similarity 71.7%; Pred. No. 7,8e-57;
	Matches 231; Conservative 41; Mismatches 46; Indels 4; Gaps 2;
QY	9 PPAISQDRAFDALPLDPAMVAQCPOVNIINTKLGYNNAYCAGTGVIDPNGVLTN 68
Db	37 PSTALDRFNRPFLPLNPAAMVA--PPQVNISTRGLGNSAVAGTGIVIDDSSGVLTN 93
QY	69 NHVIAGATDNAPSVSGSQTYGVDDVGXYDRTDQVAVLQLRGAGGLPSAAICGGVAVGEPV 128
Db	94 NHVISGATDISADFGVGNKTYGVDDVGXYDRTDQVAVLQLRGASNLPTAVIGGDVAIGEPI 153
QY	129 VAMNSGGCGGTTRAVPGRVVALGQTVOASDSLTSGETLNGLIQFDAAIQPDAGGPVV 188
Db	154 VALGNTGGGGLFSVLPGRVVALNQTVQASEPLTGAQETLSGLIQVDAPIKPGDSGGPW 213
QY	189 NLGQVVVMGMTAASNDFOLSGGGGFPAIPGQAIAIQAQRSGGSPTVHIGPTAFGLG 248
Db	214 NSRGOVVGMNTATNDNYKM-LGGCFPAIPGQAEVVGARSAGSNVTVHIGPTAFFGLG 272
QY	249 VVDNNGGARVQRVVGSAASAISLGISTGVNITAVDGPAINSATAMADALNHHPGDVTSV 308
Db	273 VLDNNGGARVARVATGPAAAGISVGDIIITSVDGVPISBATMTNVLPVPHPGETVAV 332
QY	309 TWQTKSGTGTGNTVLAEGPPA 330
Db	333 NYSAGGCDUTANVTLAEGPPA 354
RESULT 5	
Q8VK4A	PRELIMINARY; PRT; 446 AA.
ID	Q8VK4A
AC	Q8VK4A; 2002 (TrEMBLrel. 20, Created)
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Heat shock protein HtrA, putative.
GN	Mt101.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RX	NCBI_TaxId=1773;
FN	[1]
RP	SEQUENCE FROM N.A.
RC	SRAIN=CDC 1551 / Oshkosh;
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA	Peterson J., DeBoy K., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA	Kolony J.F., Nelson W.C., Unayai L.A., Ermolaeva M.D., Salzberg S.L.,
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA	Biswal W.;
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
RL	Submitted (Apr-2003) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AE006995; AAA45259.1; -.
DR	TIGR; MT1011; -.
DR	GO; GO:0004295; F:trypsin activity; IEA.
DR	GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	InterPro; IPR009003; Cys_Ser_trypsin.
DR	InterPro; IPR001478; PDZ.
DR	InterPro; IPR001254; Peptidase-S1.
DR	InterPro; IPR001940; Peptidase_S1C.
DR	Pfam; PF00595; PDZ; 1.
DR	Pfam; PF00089; trypsin; 1.

```

DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SMO0228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
SQ SEQUENCE 446 AA; 44484 MW; 54170CBEA8FE72B CRC64;

Query Match
Best Local Similarity 37.1%; Score 457.5; DB 16; Length 446;
Matches 121; Conservative 56; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGQVNNINIKLGVNNAVGAGTGIVDPNGVLTNNHVIAGAT- 76
Db 129 PAANMPGSGVEQAAKVPVSVVWMLTDLGRQSE--EGSGIILSAEGLIITNNHVIATAAK 186
QY 77 -----DINAFVSGSQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIGGG--VAVGEP 127
Db 187 PPLGSPPKTTVTFSQRTAPFTVVGADPTSDIAVVRVQVSGLTPTISLSSSDDLVRVGP 246
QY 128 VVAMNGSGGCGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPDAGGPV 187
Db 247 VLAIGSPGLGEGT--VTGTIVSALNRPVSTTGE--AGNQNTVLDAIQTDAAINPGNSGAL 303
QY 188 VNLGLGVVGNMTA---ASDNFQLSQG--GCGFAIPIGQAMAIAGQIRSGGSPVTHIGP 241
Db 304 VVNAQLGVNSAIATLGADSDAQSGSIGLGFPAIPVDQAKRIADELISGTGA-----S 357
QY 242 TAFILGLGV--DNNNGGARVORVVGSAASLGISTGDTVITAVDGAIPNSATAMADALNGH 300
Db 358 HASLGLGVQVTKDTPGAKIVVVAGGAANAAGVPKGVVTVKVDPRPINSADALVAARSK 417
QY 301 HPGDVISVTWQKSGGTRTGNVTIAE 326
Db 418 APGATVALTFQDPGSGSRTVQVTLGK 443

RESULT 6
QYU0X2 PRELIMINARY; PRT; 464 AA.
AC QYU0X2; 2003 (TEmBLrel. 25, Created)
DT 01-OCT-2003 (TEmBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TEmBLrel. 25, Last annotation update)
DE Probable serine protease (Serine proteinase) (EC 3.4.21.-).
GN MB1009.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AP2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garner T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchon S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
RA "The complete genome sequence of Mycobacterium bovis."
RA Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL EMBL; BX248337; CAD93870.1; -.
DR Hydrolase; Complete proteome.
SQ SEQUENCE 464 AA; 46436 MW; AE93A4BB3FFA9BE3 CRC64;

Query Match
Best Local Similarity 37.1%; Score 457.5; DB 16; Length 446;
Matches 121; Conservative 56; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGQVNNINIKLGVNNAVGAGTGIVDPNGVLTNNHVIAGAT- 76
Db 147 PAANMPGSGVEQAAKVPVSVVWMLTDLGRQSE--EGSGIILSAEGLIITNNHVIATAAK 204
QY 77 -----DINAFVSGSQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIGGG--VAVGEP 127
Db 205 PPLGSPPKTTVTFSQRTAPFTVVGADPTSDIAVVRVQVSGLTPTISLSSSDDLVRVGP 264

PRINTS; PRO0834; PROTEASES2C.
SMART; SMO0228; PDZ; 1.
PROSITE; PS0106; PDZ; 1.
SEQUENCE 446 AA; 44484 MW; 54170CBEA8FE72B CRC64;

Query Match
Best Local Similarity 37.1%; Score 457.5; DB 16; Length 446;
Matches 121; Conservative 56; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGQVNNINIKLGVNNAVGAGTGIVDPNGVLTNNHVIAGAT- 76
Db 147 PAANMPGSGVEQAAKVPVSVVWMLTDLGRQSE--EGSGIILSAEGLIITNNHVIATAAK 204
QY 77 -----DINAFVSGSQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIGGG--VAVGEP 127
Db 205 PPLGSPPKTTVTFSQRTAPFTVVGADPTSDIAVVRVQVSGLTPTISLSSSDDLVRVGP 264

```


GO; GO:000508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR PRINTS; PR00889; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Protease; Complete proteome.
 SQ SEQUENCE 375 AA; 39724 MW; 0D2602EB534915F5 CRC64;

Query Match 23.1%; Score 393; DB 16; Length 375;
 Best Local Similarity 33.6%; Pred. No. 1e-13;
 Matches 108; Conservative 50; Mismatches 113; Indels 50; Gaps 10;

QY 10 PALSQDRPADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGTGIVDPNGVLTN 68
 Db 67 PILSPDFPQFPFGALPQE-----DRLRGQSGFIDPSGIWVN 108
 QY 69 NHVIGATDINAFSVSGQTYGVVGVDRDTPQVAVLQRLG-AGGLPSAAIGGG--VAVG 125
 Db 109 AHVVSQADTAVY-RLKDGRFVEGVRGVDEVSGLAIVKLKGVTEPLTAPLGDSSEVKVG 167
 QY 126 EPVWAGNSGGGGTTPRAVGRVVALG--QTVAQSDSLTGAEETLNGLIQDAAIQPGDA 183
 Db 168 DWAIAGNPLGDNLT-----VTLGIIISLHRSQAQVGPDKLDFIQTDAAINPNS 219
 QY 184 GGPVNVGLGVVGMNTAASDNFQSQGGQGFAPICQAMAIAGQIRSGGSGTWHIG--P 241
 Db 220 GGPLLNEAGEVIGINTA----IRADAMGIGFALPINKAKALQARLIRGEKIQHAVIGQM 275
 QY 242 TAFGLGVVDNNG-----NGARVQRVVGSAAPASLGISTGDIVTAVDGAPINSA 290
 Db 276 TTTFTPAKKNANPNPSPILPEVANGVLVQLVNTPAKAGLRGWDVITAVDGPITSA 335
 QY 291 TAVDALNGHFGDVISVTWQ 311
 Db 336 DQLQTIIVDSAAVGVNLTVQ 356

RESULT 12

Q9FBK9 PRELIMINARY; PRT; 542 AA.
 ID Q9FBK9
 AC Q9FBK9
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC05149 OR SC08.12.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D., Harris D.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8943436;
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

Mol. Microbiol. 21:77-96(1996).
 [4]
 RL SEQUENCE FROM N.A.
 RN STRAIN=A3(2) / M145;
 RC MEDLINE=2196410; PubMed=12000953;
 RX Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Gobie A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RA "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RT Nature 417:141-147(2002).
 RL EMBL; AL939122; CAC01350.1;
 DR GO; GO:0008233; P:peptidase activity; IEA.
 DR GO; GO:0004295; P:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00889; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;

Query Match 23.0%; Score 390.5; DB 16; Length 542;
 Best Local Similarity 32.3%; Pred. No. 2.2e-13;
 Matches 111; Conservative 56; Mismatches 132; Indels 45; Gaps 11;

QY 9 PPALSQDRPADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGTGIVDPNGVLTN 68
 Db 217 PEAAERD-----PDSVAGIAARALPSVVTLH--VSGEAGTGTGFLVDGRGHLTN 266
 QY 69 NHVI--AGATDINAFSVSGQTYGVVGVDRDTPQVAVLQRLGAGGLPSAAIGG--GVAV 124
 Db 267 NHVVEPAGSGEITVTFTNSGDTAEAEVVGDRDSGYDLAVVKVGTGLTPLPLGNSDNVRV 326
 QY 125 GEPVNVANGSGGGTTPRAVGRVVALGQTVQA--SDSLTGAEETLNGLIQDAAIQPGDA 183
 Db 327 GDPVVAIGAPFDLAGT--VTSGIISAKERPITAGGEGSGDSISYVDALQTDAPINPNS 384
 QY 184 GGPVNVGLGVVGMNTA-----ASDNFQSQGGQGFAPICQAMAIAGQIRSGGGS 234
 Db 385 GGPLLDARGAIGINSAIRSADSGSTESDDQAGSIGLGFAPINQGRVAEEL----- 438
 QY 235 PTVHIGTAFGLGV-VDNN--NGARVQ-----RVVGSAPASLGISTGDIVTAVD 283
 Db 439 --INTGKAHPFVIGITLDMNYTGDGARVSARKGSGDGPVTTGGGKAGIKPGDIVTAVD 496
 QY 284 GAPINSATAMADALNGHHPGDVISVTWQTKSGGTRTGNVTIAEG 327
 Db 497 GQRVHSGEELIVKTRAPRPGDRLELTLDQDKETKVSILVLSGSG 540

RESULT 13
 Q82FM9 PRELIMINARY; PRT; 472 AA.
 ID Q82FM9
 AC Q82FM9
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SAV4223.
 OS Streptomyces avermitilis.

```

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005038; BAC7195.1; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASIS2C.
DR Protease; Complete proteome.
SQ SEQUENCE 472 AA; 46086 MW; AE04FAA409A3B7E0 CRC64;

Query Match 22.4%; Score 381; DB 16; Length 472;
Best Local Similarity 32.2%; Pred. No. 6.1e-13;
Matches 102; Conservative 59; Mismatches 133; Indels 24; Gaps 10;

QY 28 SAMVAQGVQVNVNINIKLGYNNVAGTGTIVDPNGVLTNNHVIAGATDINAFSV--GS 85
DB 157 AGVAAKALPTVTIEAQSSGEG-GTGTGFVFDKQGHIVTNHVAEAVDGGKLTATFPD 215
QY 86 GQTVGVVGVYDRTQDVAVLQLRGA--GGLPSAAIGGG--VAVGEPVAVMNGSGGGGTPT 142
DB 216 GKYNAEVVGHAQGYDVAVVVKLNKAPSDLQPLTGLDPSDKVAVGDSIAIGAPFLSNT-- 273
QY 143 AVGRVVALQQTVAQSDSLTGAETLNGLIQFDAAIQPDGAGPVNGLGVVGMNTA-- 200
DB 274 VTTGIISAKRPVASSDSSSSASTYSALQTDASINPNSGGPLDQAQSSVIGINSATIQ 333
QY 201 ASDNFOLSGGQ-----GFAIPGOMATAGAIRSGGSPVTHIGPTAFL-----GLGV 249
DB 334 SSSSGGLSGSGSGSGLGLFAIPINQAKYVAQELIKTPVPVIGASVLSGEGTGAKI 393
QY 250 VDNNGGARVQRVVGAPAAASLGISTGDVITAVDGPAPINSATAMADALNGHPGDVISVT 309
DB 394 TEGASGS--DAITPNPAAKAGLKPQDVIITKLDMDVIDSGPTLIGEIWTHRPGATVKLT 451
QY 310 WQTKSGSTRTGNVTLAE 326
DB 452 Y-TRDGKARTDVLIGE 467

RESULT 14
Q82IL8 PRELIMINARY; PRT; 619 AA.
AC Q82IL8;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)

```

```

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative serine proteinase.
GN SAV3115.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005033; BAC70826.1; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASIS2C.
DR SMART; SM00228; PDZ; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 619 AA; 64251 MW; D2D3C72A67D3B3A6 CRC64;

Query Match 22.4%; Score 381; DB 16; Length 619;
Best Local Similarity 33.7%; Pred. No. 8.4e-13;
Matches 114; Conservative 54; Mismatches 126; Indels 44; Gaps 12;

QY 13 SQDRFADFPALPLDPSAMVAQGVQVNVNINIKLGYNNVAGTGTIVDPNGVLTNNHVI 72
DB 298 SEERAADSV-----GIAARALPSVVTLVHK--GSAAGTGTGTVLDGRGHILTNHVV 349
QY 73 --AGATDINAFSVGSGQTVGVYDRTQDVAVLQLRGAAGLPSAAIGG--GVAVGEPV 128
DB 350 EPAGSSGEISVTFSGGTAKATVVGROSGYDLAVVKVSGVGLKMPPLGNSDNNVQGDVP 409
QY 129 VAMNGSGGGGTTPRAVGRVVALQQTVA--SDSLTGAETLNGLIQFDAAIQPDGAGPV 187
DB 410 VAIGAPFLANT--VTSGIISAKERPITAGGKGDSDVSVDALQTDAPINPNSGGPL 467
QY 188 VNLGVQVVGWMTAA-----SNFQLSGGQ-----GFAIPGOMATAGAIRSGGSPVTHIG 240
DB 468 LDSKRAVVGINSATRSADSSDQSGAGSITGLGFAIPVNOAKRVAAELINTGATHPEVIG 527
QY 241 PTAFLGLGVVNNNGNGARV-----QVVVGSAPAAASLGISTGDVITAVDGPAPINSATA 292
DB 528 VILDM-----DVTGDGARVGTKNDSGSPVTRGPGDRAIGIAGDVTITEVDGQRIHSSEE 582
QY 293 MADALNGHPGDVISVTWQTKSGGTRTGN---VTLAEG 327
DB 583 LIVKIRAHPRGDRALATVE-----RDGKEKPVTLVLG 614

RESULT 15
Q8FRI7

```

ID	Q8FR17	PRELIMINARY;	PRT;	473 AA.
AC	Q8FR17;			
DT	01-MAR-2003	(TrEMBLrel. 23, Created)		
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)		
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
DE	Putative serine protease, heat shock protein.			
GN	C80950.			
OS	Corynebacterium efficiens.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Corynebacteriaceae; Corynebacterium.			
OX	NCBI_TaxID=152794;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;			
RA	Kawarayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,			
RA	Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,			
RA	Usuda Y., Sugimoto S.;			
RT	"The entire genomic sequence of Corynebacterium efficiens YS-314.";			
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AP005217; BAC17760.1; -			
DR	GO; GO:0003773; F:heat shock protein activity; IEA.			
DR	GO; GO:0008233; F:peptidase activity; IEA.			
DR	GO; GO:0004295; F:trypsin activity; IEA.			
DR	GO; GO:0007242; P:intracellular signaling cascade; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR009003; Cys_Ser_trypsin.			
DR	InterPro; IPR001478; PDZ_1.			
DR	InterPro; IPR001254; Peptidase_S1.			
DR	InterPro; IPR001940; Peptidase_S1C.			
DR	Pfam; PF00595; PDZ; 1.			
DR	Pfam; PF00889; trypsin; 1.			
DR	PRINTS; PR00834; PROTEASES2C.			
DR	SMART; SM00228; PDZ; 1.			
DR	PROSITE; PS50106; PDZ; 1.			
KW	Heat shock; Protease; Complete proteome.			
SQ	SEQUENCE 473 AA; 47757 MW; 5006C9561F8C524 CRC64;			
Query Match 22.1%; Score 376; DB 16; Length 473;				
Best Local Similarity 34.0%; Pred. No. 1.1e-12;				
Matches 107; Conservative 51; Mismatches 131; Indels 26; Gaps 9;				
QY	22	ALPLDPSMVAQVGPQVNVINTKLGYNNAVAGTGIVIDPENGVLTNHVIAGATDINAF	81	
Db	170	AEPGSAEQVADQVLPVSVISQALT--RNSASEGSGSISSDGFVMTNNHVAQVEQNGVL	227	
QY	82	SV--GSGQTVGVGVVDRVDVAVLQLRGAGGLPSAAIG--GGVAVGEPVWAMGNSGQ	137	
Db	228	QVTMSDGSSEADFIAGDVSTDIATIKRNATDLPNINFSDSLAVGQEVLAQVSPGL	287	
QY	138	GGTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDAIQFDAGGFVWNLGQVVGCM	197	
Db	288	SST--VTSGIVSALNRPVRAAGD--GGESSLIDAIQTDAAINPNSGGFLVDMEGNLICM	343	
QY	198	NTPAADNFQLSQ-----GQGFATPGQAMAIAGQIRSGGSPVTHIGTAFGLGVV	250	
Db	344	NSVIA---SISSTGDTGGSIGLGFAPANFAKRVATQLTETGVATQPMIGVTLGANY-	399	
QY	251	DNNGNGARVQRVVGSAAPASLSIGTSDVITAVDGPAINSATAMADALNGHFGPDVISVTW	310	
Db	400	-----SCALLIASVQDGGFGAAGLKPQGVVMTLNDVRVIDSADALIAAVRSQDGEIVTLTV	455	
QY	311	QTK--SGGTGTGNVTL	324	
Db	456	TAQDTSQTRQVEVTL	470	

Search completed: June 22, 2004, 17:22:47
Job time : 27.9826 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 9.233 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-6
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSDQRFADF.....QTKSGGTRTGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:.*
1: Pir1:.*
2: Pir2:.*
3: Pir3:.*
4: Pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1642	96.7	355	2 F70983	probable serine pr
2	1228.5	72.3	361	2 S47170	hypothetical prote
3	1195	70.4	354	2 A87242	probable secreted
4	456.5	26.9	464	2 C70821	probable serine pr
5	447	26.3	382	2 H86930	probable secreted
6	447	26.3	452	2 T45448	probable serine pr
7	371	21.8	394	2 S74643	proteinase hhoA (E
8	365	21.5	407	2 AG2150	serine proteinase
9	350	20.6	433	2 H97199	htrA-like serine p
10	349	20.6	452	2 S77538	serine proteinase
11	335	19.7	514	2 A82581	periplasmic protei
12	332.5	19.6	362	2 T35287	probable secreted
13	330	19.4	416	2 AB2057	serine proteinase
14	329	19.4	398	2 B71284	probable periplasm
15	329	19.4	441	2 E75357	probable periplasm
16	326.5	19.2	429	2 AD1894	serine proteinase
17	316.5	18.6	474	2 F83550	serine proteinase
18	316	18.6	355	1 JC6052	trypsin-like prote
19	316	18.6	355	2 D91142	trypsin-like prote
20	315	18.6	355	2 G85987	trypsin-like prote
21	314.5	18.5	401	2 AD2451	serine proteinase
22	312	18.4	408	2 H86891	exported serine pr
23	310.5	18.3	352	2 E82307	trypsin-like prote
24	309.5	18.2	455	2 AB0909	serine protease (E
25	309	18.2	416	2 S75445	proteinase hhoB (E
26	307.5	18.1	348	2 H96956	serine protease Do
27	307.5	18.1	455	2 C91142	serine endoprotein
28	306	18.0	455	2 F85987	serine endoprotein
29	305.5	18.0	513	2 AD3418	proteinase DO (EC

30 304.5 17.9 513 2 140060 serine proteinase
31 303.5 17.9 474 1 140059 htrA-like protein
32 303.5 17.9 474 2 AI3349 proteinase DO (EC
33 302.5 17.8 455 2 JC6051 trypsin-like prote
34 301 17.7 328 1 D69109 serine proteinase
35 299.5 17.6 356 2 AC0909 trypsin-like prote
36 298 17.6 459 2 F72359 periplasmic serine
37 296 17.4 457 2 AG0433 proteinase (EC 3.4
38 295 17.4 499 2 B81914 probable periplasm
39 293 17.3 475 1 S15337 heat shock protein
40 291.5 17.2 363 2 C87336 serine proteinase
41 291.5 17.2 449 2 A69643 serine proteinase
42 291.5 17.2 474 2 S45229 proteinase DO (EC
43 291.5 17.2 474 2 B85500 proteinase DO (EC
44 291.5 17.2 474 2 E90649 proteinase DO (EC
45 291 17.1 476 2 H71936 proteinase DO - He

ALIGNMENTS

RESULT 1

F70983
probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: F70983
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:19634230
A:Accession: F70983
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-355 <COL>
A:Cross-references: GB:296071; GB:AL123456; NID:93242254; PIDN:CAB09453.1; PID:92181967
A:Experimental source: strain H37RV
C:Genetics:
C:Gene: pepA
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp:

Query Match	96.7%	Score 1642;	DB 2;	Length 355;
Best Local Similarity	99.7%	Pred No. 7.8e-90;		
Matches 322;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
QY	8	APPALSDQRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGCTGVIDPNGVLT	67	
Db	33	APPALSDQRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGCTGVIDPNGVLT	92	
QY	68	NNHVIAGATDINASVSGGQTYGVYDVYDRTQDVAVQLRGAGLPSAALGGVAVGEP	127	
Db	93	NNHVIAGATDINASVSGGQTYGVYDVYDRTQDVAVQLRGAGLPSAALGGVAVGEP	152	
QY	128	VVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPDAGGPV	187	
Db	153	VVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPDAGGPV	212	
QY	188	VNLGQVVGMMNTAASDNFQSQGGQFALPIGQAMATAGQIRSGGSGPTVHIGTAFGL	247	
Db	213	VNLGQVVGMMNTAASDNFQSQGGQFALPIGQAMATAGQIRSGGSGPTVHIGTAFGL	272	
QY	248	GVDDNNGGARVQRVVGSAASLGISTGDTVAVDGPAINSATAMADALNGHHPGVIS	307	
Db	273	GVDDNNGGARVQRVVGSAASLGISTGDTVAVDGPAINSATAMADALNGHHPGVIS	332	
QY	308	VTWTKSGGTRTGNVTLAEGPPA	330	
Db	333	VTWTKSGGTRTGNVTLAEGPPA	355	

Db 436 APGATVALTFQDSPGSGSRVQVTLGK 461

RESULT 5
H86930
probable secreted serine proteinase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C;Accession: H86930
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H86930
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-382 <STO>
A;Cross-references: GB:AL450380; NID:gl3092536; PIDN:CAC29684.1; GSPDB:GN00147
C;Genetics:
C;Gene: Mtu0176
C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 26.3%; Score 447; DB 2; Length 382;
Best Local Similarity 37.2%; Pred. No. 1.8e-19;
Matches 125; Conservative 53; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALDPSAMVAQVG-----PQVNVNITKLYNNVAGAGTGIVIDPENGVLNNHVA 73
DB 59 ASVPAANM-PSGSVEQVAVKVPVVMLETLGRQSE--EGSGVILSADGILTNHVA 115
QY 74 GA-----TDINAFSVGSGQTYGVVDVVDYDTQVAVLQLRGAGLPSAAIG 119
DB 116 VAAKPGGGGGLSPKTVTFP---DGRSTFTVVGADPTSDIAVVRVOSISGLTPTMG 172
QY 120 GG--VAVGEPVVMNGSGGQGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAA 177
DB 173 SSADLRVGPVAVVAGSPLGLAGT--VTSGIVSALNRPVSTTGE--SGNQNTVLDAQTDA 229
QY 178 IQPGDAGGPVNVGLGVGMNTAA-----SDNFQLSQGGGFAIPIGQMAIAGQIRSG 231
DB 230 INFGSGGALVNMGGVLGVNSAIFLGDAGSDGQSGSIGLGFALPVDAQRIADELIST 289
QY 232 GGSPTVHIGPTAFLGLGVVDNNGN--GARVQRVVGSAFASLIGSTGDTAVDGAIPNSA 290
DB 290 G--KATH-----ASLGVQVATDKGTPGAKVMDVVAGGAANAAPKGVVLTKVDDRLISSA 343
QY 291 TAMADALNGHHPGDVISVTWTKSGGTRTGNVTLAE 326
DB 344 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 379

RESULT 6
T45448
probable serine proteinase (EC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Dec-2002
C;Accession: T45448
R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: D22967
A;Accession: T45448
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-452 <JAM>
A;Cross-references: EMBL:AL035500; PIDN:CAB36690.1
A;Experimental source: cosmid L373
C;Genetics:
C;Note: MLCB373.28
C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

F1182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 26.3%; Score 447; DB 2; Length 452;
Best Local Similarity 37.2%; Pred. No. 2.2e-19;
Matches 125; Conservative 53; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALDPSAMVAQVG-----PQVNVNITKLYNNVAGAGTGIVIDPENGVLNNHVA 73
DB 129 ASVPAANM-PSGSVEQVAVKVPVVMLETLGRQSE--EGSGVILSADGILTNHVA 185
QY 74 GA-----TDINAFSVGSGQTYGVVDVVDYDTQVAVLQLRGAGLPSAAIG 119
DB 196 VAAKPGGGGGLSPKTVTFP---DGRSTFTVVGADPTSDIAVVRVOSISGLTPTMG 242
QY 120 GG--VAVGEPVVMNGSGGQGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAA 177
DB 243 SSADLRVGPVAVVAGSPLGLAGT--VTSGIVSALNRPVSTTGE--SGNQNTVLDAQTDA 299
QY 178 IQPGDAGGPVNVGLGVGMNTAA-----SDNFQLSQGGGFAIPIGQMAIAGQIRSG 231
DB 300 INFGSGGALVNMGGVLGVNSAIFLGDAGSDGQSGSIGLGFALPVDAQRIADELIST 359
QY 232 GGSPTVHIGPTAFLGLGVVDNNGN--GARVQRVVGSAFASLIGSTGDTAVDGAIPNSA 290
DB 360 G--KATH-----ASLGVQVATDKGTPGAKVMDVVAGGAANAAPKGVVLTKVDDRLISSA 413
QY 291 TAMADALNGHHPGDVISVTWTKSGGTRTGNVTLAE 326
DB 414 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 449

RESULT 7
S74643
proteinase rhoA (EC 3.4.-.-) - Synecocystis sp. (strain PCC 6803)
N;Alternate names: protein all1679
C;Species: Synecocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
C;Accession: S74643
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
S.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S74643
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-394 <KAN>
A;Cross-references: EMBL:D90900; GB:AB001339; NID:gl651769; PIDN:BAAL6795.1; PID:gl65186f
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: rhoA
C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypt
C;Keywords: hydrolase; proteinase

Query Match 21.8%; Score 371; DB 2; Length 394;
Best Local Similarity 31.3%; Pred. No. 5.7e-15;
Matches 117; Conservative 50; Mismatches 111; Indels 96; Gaps 15;

QY 9 PPA-----LSQDRFADFALPLDPSMVAQVQVNVNITK----- 44
DB 39 PPAVITAQASVPLTSSFV-----AAVSRSGFAVVRIDTETVTRTDPILDDP 89
QY 45 -----LGVNNAV-----GAGTGIVIDPENGVLNNHVIAGATDNAPSVSGSQTYGV 91
DB 90 PFQEFFGRSPVPPRERRIAGQSGFIIDNSGIILTNHAVVDGASKV-VVTLRDRGRFDG 148
QY 92 DVGYDRTQDVAVLQLRGAG-GLPSAAIG--GGVAVGEPVVMNGSGGQGTTPRAVGRV 148
DB 149 QVRGTDEVTDLAVVKIEPQGSALFPVPLGTSSNLQVGDMAIAGVNPGLONT----- 200
QY 149 VALG--QTVQASDLSLTGAETLNGLIQFDAAIQDGDAGGPVNVGLGVGMNTAASDNFQ 206

Db 201 VTIGIISTLGRSAQAQIPKRVFEITQDAINPNSGGFLNARGEVIGINTA---IR 256
 QY 207 LSGGGGFAIPQQAIAQIRSGGSPVTHGPTAFGLGV---VNNGN----- 255
 Db 257 ADATGIGFAIPQQAIAQNTLAAGTVEHPYIG-VQMNITVDQAQQNRNPNPFFIIP 315
 QY 256 ---GASVQVVVGSAASGISGIVTAVDGAIPNSATAM-----ADALN 298
 Db 316 EVGILVMRVLPTPAERAGIRGDIIVAVDGPISDGAIRQIVGQAGLNKALKLDLLR 375
 QY 299 GHPGDIIVSVTQOT 312
 Db 376 GDRR---LSITVQT 386

RESULT 8
 AG2150
 serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AG2150
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, S.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AG2150
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-407 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA74457.1; PID:ga7131851; GSPDB:GN00179
 C:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr2758
 C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 21.5%; Score 365; DB 2; Length 407;
 Best Local Similarity 31.2%; Pred. No. 1.3e-14;
 Matches 109; Conservative 52; Mismatches 118; Indels 70; Gaps 11;
 QY 28 SAMVAQVGPQVNVNKLGYNNAV-----GAGTGIVD 60
 Db 74 TAAVNRVGSAAVRIDTERTTRVDPLEDFPRFRFEGFQGLPPEQMRGLGSGFIID 133
 QY 61 PNGVLTNNHVIAGATDINAFVSGGQTYGVYDRTQDVAVLQLRGAGGLPSAAIG- 119
 Db 134 KSGILITNAHVVDKADRTV-RLKDGSRFDKQGVIDEVTDLAVVKINAGNSLFVAPLGS 192
 QY 120 -GGVAGPEPVVAMNSGGGQGTTPRAGRVVAILG-QTVOASDLSLTGAETLNGLIQFPA 176
 Db 193 SNNVQVGDMAIAGVNPGLGPDNT-----VTILGIVSTLKRSAQVIGITDKLDFIQTPA 244
 QY 177 AIOPGDAGGPVAVNGLGVGVGMNTAASDNFOLSGGQGFAPITQQAIAQIRSGGSP 236
 Db 245 AINPNSGGFLNDKKEVIGINTA---IFADWAGIGFPAIPIDKAKIATQLERDQ--KV 298
 QY 237 VHIGPTAFGLGV-----VDNNG-----NGARVORVVGSAAPASLSTGSD 277
 Db 299 AH-----PYLGVMQATLTPELAAQNNIDPNSAFAPPEVNGVLYVIRVVPNSPAAAGIRGD 354
 QY 278 VITAVDGAIPNSATAMADALNGHPGDIVSVTQTKSGGTRTGNVTTLAE 326
 Db 355 VILQVDCAITTAELQNVNNSRELGAQLQVRLQ-RGNQTQQLSVRTAE 402

RESULT 9
 htrA-like serine protease (with PDZ domain) [imported] - Clostridium acetobutylicum
 H97199
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Dec-2002
 C:Accession: H97199

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R. J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: H97199
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-433 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK80387.1; PID:gl5025449; GSPDB:GN00168
 C:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC2433
 C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 20.6%; Score 350; DB 2; Length 433;
 Best Local Similarity 31.6%; Pred. No. 1.1e-13;
 Matches 100; Conservative 51; Mismatches 111; Indels 54; Gaps 10;
 QY 25 LDPSAMVAQVGPQVNVNKL-----GYNNVAGTGTVIDPNGVLT 67
 Db 122 LTVSQIVKVKVSPAVVGVTTKTTVTQNDFFSGSSNGSGSTQEGMSGIIFNNDGYILT 181
 QY 68 NNHVIAGATDINAFVSGGQTYGVYDRTQDVAVLQLRGAGGLPSAAIGGVA--V 124
 Db 182 NYHVIKGAOKT-AVILNNKKEVSAKVNYDEANDIAVKITGFTVPGVAELSSASLNV 240
 QY 125 GEPVAMNSGGGQGTTPRAGRVVAILGQTVQASDLSLTGAETLNGLIQFPAIQPDAG 184
 Db 241 GDSVVAIGNPLGKEFLGTVTGTVVSAVNRVAVSE---GQKQT---YIQTDAAINPNSG 294
 QY 185 GPVNGVLGVGMNTAASDNFOLSGGQ---GFAIPITQQAIAQIRSGGSPVTHIGP 241
 Db 295 GPLVNSFGQVVGINS-----KISENGVEGIGSFIPIDTVKSKIQNLSK-----P 339
 QY 242 TAFILGLG--VVD-----NNGNGARVQVRVVGSAAPASLGTIGDVTITAVDGAIPNSATA 292
 Db 340 ILMGLISGEAVDKSTASQHNIPQGVYIEQIQDFSSAQKAGMVGVDVITKFDGKKVTS 399
 QY 293 MADALNGHPGDIVSV 308
 Db 400 IDSIRKSHNSGDTVQV 415

RESULT 10
 S77538
 serine proteinase (EC 3.4.21.-) htrA - Synechocystis sp. (strain PCC 6803)
 N:Alternate names: protein slr1204
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
 C:Accession: S77538
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; K. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M. DNA Res. 3, 103-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S77538
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-452 <KAN>
 A:Cross-references: EMBL:D90905; GB:AB001339; NID:gl523360; PIDN:BAAL7385.1; PID:gl5246;
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: htrA
 C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 20.6%; Score 349; DB 2; Length 452;
 Best Local Similarity 29.9%; Pred. No. 1.3e-13;
 Matches 109; Conservative 60; Mismatches 113; Indels 82; Gaps 15;

```
QY 24 PLDPAMVAQV---GPOVNTNTKLGYNNAV-----G 52
Db 112 PREPNFVVDVVESTGPAVRINAQTVKSPQAFNDPFLQFPFGSQMPMPNERVORG 171
QY 53 AGTVIDPNDVNLNNHVIAGATINAFSGSGQTVGVVYDRTQDVAVLQRLGAGG 112
Db 172 TSGSGFVSDGSKIPTNAHVVDGADETV--TLKDGSRFPGRVMSGSDPSTDVAVKIE--AGD 229
QY 113 LPSAAGGG--VAVEGPVVMGNSGGQGTTPRAVGRVVALGQTVQASDLSLTGABETLNG 170
Db 230 LFTVALGSDHDLQGEWAIRAGNPLGLDNT--VTGILSATGR--RSAD--IGVPDKRVE 283
QY 171 LIQFDAAIOPDAGGVVNGLVGMNTAASDNFQLSGGQGAIPITGOAMAIAGQIRS 230
Db 284 FIOTDAINPNSGGFLLNADQVIGMNTAIONAQ-----GIGFAIPINKAQEIAQOLIA 339
QY 231 GGSSTVHIGPTAFILGLGVWNN-----NGGARVQRVVSAPASLGI 273
Db 340 TG--KVHE-----AYLGIQNVTVTPELQSIQRETGNIPVDKGVVIMQVMPNSPAIAKL 393
QY 274 STGDVITAVDGAIPINSATAMADALNGHHPGDVISVTWTKSGGTRTG---NVTLAG--- 327
Db 394 EOGDVLQSLQGGQPVENAEQVSLVGLKLVGDEVEL-----GILNGQQQLNTVTIGALP 447
QY 328 --PP 329
Db 448 SAPP 451

RESULT 11
A:Accession: A82581
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <SIM>
A:Cross-references: GB:AE004037; GB:AE003849; NID:99107394; PIDN:AAF95040.1; GSPDB:GN001
A:Experimental source: strain 945c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2241
C:Superfamily: Helicobacter serine proteinase

Query Match 19.7%; Score 335; DB 2; Length 514;
Best Local Similarity 32.3%; Pred. No. 1e-12;
Matches 108; Conservative 49; Mismatches 107; Indels 70; Gaps 13;

QY 23 LPLDSAMVAQVGPQVNTNT-----KLGYNNAV----- 51
Db 54 LP-DFTQLVDVQVGPVNTETITRKKVGRGILNDNDIPFRFRFPDQMPNQPRG 112
QY 52 -----GAGTGIVDPNGVLTNNHVIAGATINAFSGSGQTVGVVDVYDRTQ 100
```

```
Db 113 GQDDEGGIAGKMGSGFLISKDGVILTNHVTGASEV-TIKLTDRREFKAKIIGSEQY 171
QY 101 DVAVLQRLGAGGLPSAAGGVAV--CEPVVMGNSGGQGTTPRAVGRVVALGQTVQAS 158
Db 172 DVALLKI-DANKLFTVRIGDSSSLKSGQVVALGSPGLDHSVTA--GIVSALGRS----- 224
QY 159 DLSLTGABETLNGLTQFDAAIQPDAGGVVNGLVGMNTAASDNFQLSGG--GQGFAL 216
Db 225 ---TSDQRYVPFIQTDVPIQNQNSGGFLLNTRGEVIGIN---SQIFSAGGYMGISFAL 278
QY 217 PIGOMAIAGQIRSGG-----GSPVTHIGPTAFI---GLGVVDNNGGARVQRVVSAPAA 269
Db 279 PINLAINAABQIRKTKGVQSRMLGVEIGPIDALKAQGLPLDS--RGALVNNIPPHSPAA 336
QY 270 SLGISTGDVITAVDGAIPINSATAMADALNGHHPG 303
Db 337 KAGIEVGDVIRSVNGKVISSFSDLPLPLIGNMPPG 370

RESULT 12
T35287
Probable secreted proteinase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Dec-2002
C:Accession: T35287
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21574
A:Accession: T35287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-362 <SSE>
A:Cross-references: EMBL:AL069872; PIDN:CAB51255.1; GSPDB:GN00070; SCOEDB:SC5F7.30
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5F7.30
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps
Query Match 19.6%; Score 332.5; DB 2; Length 362;
Best Local Similarity 33.6%; Pred. No. 9.6e-13;
Matches 112; Conservative 46; Mismatches 138; Indels 37; Gaps 12;

QY 8 APPALSDRADFPALPLDPSAMVAQVGPQVNTNTKLGYNNAVAGTGIVDPNGVLT 67
Db 52 APRASE-----LEADYERVIKDVLPVSVQLQA-----GDSLGSVVYDDKGHVVT 97
QY 68 NNHVIAGATINAFSGSGQTVG---VDVVGYDRTQDVAVLQL-RGAGGLPSA--AIGGG 121
Db 98 NAHVVG---DAKSPRVTTARTEGALTAKLVSSYPEQDLAVIKLDKVPKGMRAARFADSAK 154
QY 122 VAVEGPVVMGNSGGQGTTPRAVGRVVALGQTVQASDLSLTGABETLNGLIQFDAAIQPG 181
Db 155 VEVGQIVLVMGSPILGLSS--VTQGIYSATGRVTTEGSGGGGTGATIANMVQTSAAINPG 212
QY 182 DAGGVPVNGLVGMNTAASDNFQLSGG---GOGFAIPTIGQAMAIAGQIRSGG---GSP 235
Db 213 NSGALVNLQDQVIGITFLAATDPLGDSAAFGIAPASMTVTVAQIVRDGKVTDSG 272
QY 236 TVHIGPTAFILGLGVWNNNGN--GARVQRVVSAPASLGIQSTGVITAVDGAIPINSATAM 293
Db 273 RAALGITA---RTVVDSDSYRPAAGAAVVEVSDGGAADAGLRPGDVLVKLGDITDITITSL 329
QY 294 ADALNGHHPGDVISVTWTKSGGTRTGNVTLAE 326
Db 330 SEALASWRPGDRTKVTY-TRDGKHTAEVTLGE 361

RESULT 13
AB2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
```

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AB2057
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. strain PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2057
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-416 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA073707.1; PID:G17131098; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryptase
Query Match 19.4%; Score 330; DB 2; Length 416;
Best Local Similarity 28.7%; Pred. No. 1.6e-12;
Matches 100; Conservative 64; Mismatches 126; Indels 58; Gaps 12;
QY 26 DPS---AMVAQGPQVNVNINIKLYNNV-----GAGT 55
DB DNFVVGQVGVAVRIDSARTVTSRVNENFDFRFGDGVPAQPRQVRGSGS 136
QY 56 GIVDPNGVLTNNHVIAGTDINAFSGSGTYGVVGVYDRTQDVAVLQLRGAGLPS 115
DB 137 GFIISSGQLTNAHVDGDEVTV-TLKDGRFDFKVLGEDPVTDAVIQI-NANNLPT 194
QY 116 AAIQGGGVAV--GEPVAMGSGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNLGLI 173
DB 195 VAVGNSVLPQGEAIVAGNPLGNNS--VTSGIIATGCR---SGSDIGASDKRVDYIQ 248
QY 174 FDAIQDQAGGVPVNGLVGVGMVNTAASDNFOLSGGGGFAIPICQAMAIQIRSGGG 233
DB 249 TDAIINPGSGGELLNARGQVIGMVTATIQGAQ---GLGFAIPINTVQKVSQELITQK 304
QY 234 SPVHIG-PTAFELGLGVD--NNGCARVQ-----RVVGSAPAASLGISTGDEVIT 280
DB 305 VHPYLGQVQATITPQVKERINERFGRINITADRGVLLVRVPGSPAANAGLRPGDIQ 364
QY 281 AVDGAPINSATAMADALNGHHHPGDSVTSVQTKSGGTRTGNVTLARGP 328
DB 365 SINQSVTTVEQVKIVENSIGQPLQIQIE-RNGQTTQVNVSPALP 411
RESULT 14
B71284
probable periplasmic serine proteinase DO (htrA-1) - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Dec-2002
C;Accession: B71284
R;Praser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin, R.; J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDONALD, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of *Treponema pallidum*, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: B71284
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-398 <COL>
A;Cross-references: GB:AE001248; GB:AE000520; NID:G3323074; PIDN:AAC65740.1; PID:G332308
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0773
C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryptase
Query Match 19.4%; Score 329; DB 2; Length 398;
Best Local Similarity 32.1%; Pred. No. 1.7e-12;
Matches 103; Conservative 43; Mismatches 121; Indels 54; Gaps 11;
QY 29 AMVAQGPQVNVNINIKL-GYNNV-----GAGTGIVDPNGVLTNNHVIAGTDINA 80

DB 72 AVYRSANEAVNITTEVMGVNMFLEPVLPGSGSGAIIDARGVYLTNTHTVIEGASKI-Y 130
QY 81 FVSGSGQTYGVVGVYDRTQDVAVLQLRGAGLPSAAGGG-----VAVGSPVVMGNSGG 136
DB 131 LSLHDGSGYKATVGVVDRENDLAVLKFVSPGARLTVIRFGSSRNLDVGKVLAINPFG 190
QY 137 QGGTPRAVGRVVALGQTVQASDLSLTGAETLNLGLIQDAAIQDAGAGPVPVNLGVVG 196
DB 191 LART--LTVGVVSALARPIQNKSGI-----IRNMIQTDAINPNSGGPILDTQGRMIG 242
QY 197 MNTAASDNFQLSGGQGFAPICQAMAI-----GQIRSG-----GGSPTVHIGTFLGL 247
DB 243 INTVIYST-SGSSSGVGFAVPTAKRIVSELIRYGRVRKIDAEVLQVNASIAHQAQL 301
QY 248 GVDNNGNGARVQVRVGSAPAASLGIS-----TGDEVITAVDGA PIN 288
DB 302 TV---GKLLVSVQKRGSPAAQAGLRGTTAVRYGLRRAAVIYLGCDVITAINQOVA 357
QY 289 SATAMADALNGHHHPGDSVTSV 309
DB 358 NLSDYSVLEDKKPDDEVRT 378
RESULT 15
E75357
probable periplasmic serine proteinase Do - *Deinococcus radiodurans* (strain R1)
C;Species: *Deinococcus radiodurans*
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Dec-2002
C;Accession: E75357
R;White, O.; Eichen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Maitland, S.; Shen, M.; Vamathevan, J.J.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: E75357
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-441 <WHI>
A;Cross-references: GB:AE002017; GB:AE000513; NID:G6459527; PIDN:AAF11312.1; PID:G6459530
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRI756
A;Map position: 1
C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryptase
Query Match 19.4%; Score 329; DB 2; Length 441;
Best Local Similarity 29.3%; Pred. No. 1.9e-12;
Matches 106; Conservative 50; Mismatches 110; Indels 96; Gaps 11;
QY 14 QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLYNNVAVGAGTGIVDPNGVLTNNHVI 73
DB 115 QDQFGF--ALPEDOG-----GAPATGSGGFFVNAQGDIVTNNHVD 155
QY 74 GATDINAFSGSGQTYGVVGVYDRTQDVAVLQLRGAGLPSAAI-----GGVAVGE 126
DB 156 GASDITIRLHGKQTKYAKVIARAPDYDLALIR---AEGVPRASIKPLPLGDSKLDVGL 212
QY 127 PVVAMG-----NSGGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNLGLIQF 177
DB 213 KALNAGAPFNLDPSVSGIISLERQVP-----VGSREVSVQPIQTDA 256
QY 178 IQPDAGGVPVNGLVGVGMVNTAASDNFQLSGGGGFAIPICQAMAIQIRSGGSPV 237
DB 257 INPNSGGLLSSAGQVIGVNTQILTGGAGQSGAGVGAIPINTVXRLLPQLQAGK 312
QY 238 HIGTAFELGLGVDN-----NGCARVQVRVGSAPAASLG----- 272
DB 313 --GVVSPSLGVVFDLSLPPQOLKAGLPSGALLQKTYPGSPAAAGLRGNGNKL 370
QY 273 -----IST-GDEVITAVDGA PIN SATAMADALNGHHHPGDSVTSVQTKSGGTRTGNV 322

Db 371 LPSAQGTSSISTDGLITAVNGQPLEDAGSLOEAVLATGEGQPLRLTVR-RGGKTREVEV 429

QY 323 TL 324

Db 430 TL 431

Search completed: June 22, 2004, 17:24:49
Job time : 10.233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 30.5644 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-6
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSDQRFADF.....QTKSGGTRGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/UCT NEW PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/UCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1698	100.0	330	12	US-09-886-349A-6
2	1698	100.0	330	14	US-10-098-732A-6
3	1698	100.0	330	15	US-10-369-983-20
4	1698	100.0	723	15	US-10-369-983-2
5	1698	100.0	1010	15	US-10-369-983-4
6	1695	99.8	330	12	US-09-886-349A-4
7	1695	99.8	330	14	US-10-098-732A-4
8	1695	99.8	330	15	US-10-369-983-19
9	1642	96.7	355	9	US-09-712-363-161
10	1631	96.1	355	12	US-09-886-349A-2
11	1631	96.1	355	14	US-10-193-002-80
12	1631	96.1	355	14	US-10-084-843-79
13	1631	96.1	355	14	US-10-098-732A-2
14	1000.5	1016	15	US-10-369-983-18	
15	992	58.4	1154	15	US-10-369-983-16

16	991.5	58.4	825	15	US-10-369-983-14	Sequence 14, Appl
17	990.5	58.3	813	15	US-10-369-983-15	Sequence 15, Appl
18	989	58.2	1022	15	US-10-369-983-17	Sequence 17, Appl
19	987	58.1	729	12	US-09-886-349A-18	Sequence 18, Appl
20	987	58.1	729	14	US-10-098-732A-18	Sequence 18, Appl
21	987	58.1	729	15	US-10-369-983-22	Sequence 22, Appl
22	984.5	58.0	875	15	US-10-369-983-13	Sequence 13, Appl
23	984	58.0	596	9	US-09-287-849-26	Sequence 26, Appl
24	984	58.0	596	12	US-09-886-349A-20	Sequence 20, Appl
25	984	58.0	596	14	US-10-359-460-26	Sequence 26, Appl
26	984	58.0	596	14	US-10-098-732A-20	Sequence 20, Appl
27	984	58.0	729	9	US-09-287-849-2	Sequence 2, Appl
28	984	58.0	729	12	US-09-886-349A-16	Sequence 16, Appl
29	984	58.0	729	14	US-10-359-460-2	Sequence 2, Appl
30	984	58.0	729	14	US-10-098-732A-16	Sequence 16, Appl
31	984	58.0	729	15	US-10-369-983-21	Sequence 21, Appl
32	984	58.0	729	15	US-10-359-459-2	Sequence 2, Appl
33	984	58.0	930	14	US-10-098-732A-65	Sequence 65, Appl
34	984	58.0	930	15	US-10-369-983-12	Sequence 12, Appl
35	978	57.6	195	12	US-09-886-349A-8	Sequence 8, Appl
36	978	57.6	195	14	US-10-098-732A-8	Sequence 8, Appl
37	676	39.8	231	9	US-09-287-849-28	Sequence 28, Appl
38	676	39.8	231	14	US-10-359-460-28	Sequence 28, Appl
39	670	39.5	132	9	US-09-759-143-819	Sequence 819, App
40	670	39.5	132	9	US-09-759-143-848	Sequence 848, App
41	670	39.5	132	9	US-09-780-669-819	Sequence 819, App
42	670	39.5	132	9	US-09-780-669-848	Sequence 848, App
43	670	39.5	132	9	US-09-822-827-819	Sequence 819, App
44	670	39.5	132	9	US-09-822-827-848	Sequence 848, App
45	670	39.5	132	9	US-09-895-793-819	Sequence 819, App

ALIGNMENTS

RESULT 1
US-09-886-349A-6
; Sequence 6, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; TYPE: PRT
; LENGTH: 330
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Ra35FLMutSA
US-09-886-349A-6

Query Match	100.0%;	Score 1698;	DB 12;	Length 330;
Best Local Similarity	100.0%;	Pred. No. 1.5e-121;	Indels 0;	Gaps 0;
Matches 330;	Conservative 0;	Mismatches 0;		
Qy	1	MHHHHHAPPALSDQRFADFPALP	LDPSAMVAQVGPQVNVINTKLGYNNAVGAGTGIVID	60
Db	1	MHHHHHAPPALSDQRFADFPALP	LDPSAMVAQVGPQVNVINTKLGYNNAVGAGTGIVID	60
Qy	61	PNGVVLNNHVIAGATDINAFSVGSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG	120	
Db	61	PNGVVLNNHVIAGATDINAFSVGSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG	120	

QY 121 GVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
Db 121 GVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVWGLGVVGNMTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVWGLGVVGNMTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTSGDVITAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTSGDVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 2

US-10-098-732A-6

; Sequence 6, Application US/10098732A

; Publication No. US20030175294A1

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Brannon, Mark

; APPLICANT: Guderian, Jeffrey

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a

; TITLE OF INVENTION: Leishmania Antigen

; FILE REFERENCE: 014058-012010US

; CURRENT APPLICATION NUMBER: US/10/098, 732A

; CURRENT FILING DATE: 2003-04-29

; PRIOR APPLICATION NUMBER: US 60/275,837

; PRIOR FILING DATE: 2001-03-13

; NUMBER OF SEQ ID NOS: 80

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 6

; LENGTH: 330

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA

US-10-098-732A-6

Query Match 100.0%; Score 1698; DB 14; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.5e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNVAVGAGTGIVID 60
Db 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNVAVGAGTGIVID 60
QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGYDRTQDVAVLQIRGAGGLPSAAIGG 120
Db 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGYDRTQDVAVLQIRGAGGLPSAAIGG 120
QY 121 GVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
Db 121 GVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVWGLGVVGNMTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVWGLGVVGNMTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTSGDVITAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTSGDVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 3

US-10-369-983-20

; Sequence 20, Application US/10369983

; Publication No. US20030235593A1

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis

; FILE REFERENCE: 014058-009081US

; CURRENT APPLICATION NUMBER: US/10/369,983

; CURRENT FILING DATE: 2003-02-18

; PRIOR APPLICATION NUMBER: US 60/357,351

; PRIOR FILING DATE: 2002-02-15

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 20

; LENGTH: 330

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: mutated

; OTHER INFORMATION: MTB32AMutSA (Ra35 mutSA)

US-10-369-983-20

Query Match 100.0%; Score 1698; DB 15; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.5e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNVAVGAGTGIVID 60
Db 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNVAVGAGTGIVID 60
QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGYDRTQDVAVLQIRGAGGLPSAAIGG 120
Db 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGYDRTQDVAVLQIRGAGGLPSAAIGG 120
QY 121 GVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
Db 121 GVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVWGLGVVGNMTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVWGLGVVGNMTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTSGDVITAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTSGDVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 4

US-10-369-983-2

; Sequence 2, Application US/10369983

; Publication No. US20030235593A1

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis

; FILE REFERENCE: 014058-009081US

; CURRENT APPLICATION NUMBER: US/10/369,983

; CURRENT FILING DATE: 2003-02-18

; PRIOR APPLICATION NUMBER: US 60/357,351

; PRIOR FILING DATE: 2002-02-15

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 2

```
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB32-MTB39f fusion protein (MTB32MutSA)
US-10-369-983-2

Query Match      100.0%; Score 1698; DB 15; Length 723;
Best Local Similarity 100.0%; Pred. No. 3.9e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQGVQVNNINTKLYNNNAVAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQGVQVNNINTKLYNNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
QY 121 GVAAGPVPVAMGNSGGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
DB 121 GVAAGPVPVAMGNSGGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 121 GVAAGPVPVAMGNSGGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
DB 121 GVAAGPVPVAMGNSGGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGAIPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGAIPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 5
US-10-369-983-4
; Sequence 4, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB-102F fusion
US-10-369-983-4

Query Match      100.0%; Score 1698; DB 15; Length 1010;
Best Local Similarity 100.0%; Pred. No. 5.9e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQGVQVNNINTKLYNNNAVAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQGVQVNNINTKLYNNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
```

```
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
QY 121 GVAAGPVPVAMGNSGGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
DB 121 GVAAGPVPVAMGNSGGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGAIPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGAIPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 6
US-09-886-349A-4
; Sequence 4, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35 mature)
US-09-886-349A-4

Query Match      99.8%; Score 1695; DB 12; Length 330;
Best Local Similarity 99.7%; Pred. No. 2.6e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQGVQVNNINTKLYNNNAVAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQGVQVNNINTKLYNNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
QY 121 GVAAGPVPVAMGNSGGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
DB 121 GVAAGPVPVAMGNSGGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGAIPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGAIPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: wild-type
; OTHER INFORMATION: mature MTB32A (Ra35)
US-10-369-983-19

Query Match          99.8%; Score 1695; DB 15; Length 330;
Best Local Similarity 99.7%; Pred. No. 2.6e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 MHHHHHAPPALSDRFPADFPALPLDPSAMVAQVGPVNNINIKLGYNNNAVAGCTGIVID 60
DB      1 MHHHHHAPPALSDRFPADFPALPLDPSAMVAQVGPVNNINIKLGYNNNAVAGCTGIVID 60
QY      61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
DB      61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
QY      121 GVAVGEPVWANGSGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNLGIQFDDAIIQ 180
DB      121 GVAVGEPVWANGSGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNLGIQFDDAIIQ 180
QY      181 GDAGGPVYVNGLGQVVGMMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
DB      181 GDSGGPVYVNGLGQVVGMMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
QY      241 PTAFLGLGVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
DB      241 PTAFLGLGVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
QY      301 HPGDVISVTWTKSGGTRTGNVTILAEGLPPA 330
DB      301 HPGDVISVTWTKSGGTRTGNVTILAEGLPPA 330

RESULT 9
US-09-712-363-161
; Sequence 161, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161

US-10-098-732A-4
; Sequence 4, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35 mature
US-10-098-732A-4

Query Match          99.8%; Score 1695; DB 14; Length 330;
Best Local Similarity 99.7%; Pred. No. 2.6e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 MHHHHHAPPALSDRFPADFPALPLDPSAMVAQVGPVNNINIKLGYNNNAVAGCTGIVID 60
DB      1 MHHHHHAPPALSDRFPADFPALPLDPSAMVAQVGPVNNINIKLGYNNNAVAGCTGIVID 60
QY      61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
DB      61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
QY      121 GVAVGEPVWANGSGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNLGIQFDDAIIQ 180
DB      121 GVAVGEPVWANGSGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNLGIQFDDAIIQ 180
QY      181 GDAGGPVYVNGLGQVVGMMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
DB      181 GDSGGPVYVNGLGQVVGMMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIG 240
QY      241 PTAFLGLGVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
DB      241 PTAFLGLGVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
QY      301 HPGDVISVTWTKSGGTRTGNVTILAEGLPPA 330
DB      301 HPGDVISVTWTKSGGTRTGNVTILAEGLPPA 330

RESULT 8
US-10-369-983-19
; Sequence 19, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
```

```
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-161

Query Match
Best Local Similarity 96.7%; Score 1642; DB 9; Length 355;
Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 92
QY 68 NNHVIAGATDINAFSGSGQTVGVVGVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSGSGQTVGVVGVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 152
QY 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAALQPGDAGGPV 187
DB 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAALQPGDAGGPV 212
QY 188 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAALQPGDAGGPV 247
DB 213 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAALQPGDAGGPV 272
QY 248 GVVVNNNGGARVQVRVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVVNNNGGARVQVRVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQKSGGTRTGNVTLAEGPPA 355

RESULT 11
US-10-193-002-80
; Sequence 80, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skelky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-10-193-002-80

Query Match
Best Local Similarity 96.1%; Score 1631; DB 14; Length 355;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAALQPGDAGGPV 187
DB 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAALQPGDAGGPV 212
QY 188 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAALQPGDAGGPV 247
DB 213 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAALQPGDAGGPV 272
QY 248 GVVVNNNGGARVQVRVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVVNNNGGARVQVRVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQKSGGTRTGNVTLAEGPPA 355

RESULT 11
US-10-193-002-80
; Sequence 80, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skelky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-10-193-002-80

Query Match
Best Local Similarity 96.1%; Score 1631; DB 14; Length 355;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

QY 8 APPALSODRPFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 67
DB 33 APPALSODRPFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 187
DB 153 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 212
QY 188 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 247
DB 213 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 12

US-10-084-843-79
; Sequence 79, Application US/10084843
; Publication No. US20030143243A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.

Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-084-843-79

Query Match 96.1%; Score 1631; DB 14; Length 355;
Best Local Similarity 99.1%; Pred. No. 2.1e-116;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSODRPFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 67
DB 33 APPALSODRPFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 187
DB 153 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 212
QY 188 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 247
DB 213 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 13

US-10-098-732A-2
; Sequence 2, Application US/10098732A
; Publication No. US20030175294A1

GENERAL INFORMATION:

APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
CURRENT FILING DATE: 2003-04-29
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2

LENGTH: 355
TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: MTE32A (Ra35FL)

US-10-098-732A-2

Query Match 96.1%; Score 1631; DB 14; Length 355;
Best Local Similarity 99.1%; Pred. No. 2.1e-116;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSODRPFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 67
DB 33 APPALSODRPFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPGRVVWALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 187

Db 153 VVAMGNSGGGTPRAVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPV 212
Qy 188 VNGLGQVVGWNTAASDNFQLSGGGQFAIPICQAMAIAGQIRSGGSPTHVIGPTAFGL 247
Db 213 VNGLGQVVGWNTAASDNFQLSGGGQFAIPICQAMAIAGQIRSGGSPTHVIGPTAFGL 272
Qy 248 GVVDNNGNGARVQVVVGAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
Db 273 GVVDNNGNGARVQVVVGAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
Qy 308 VTWQKSGGTRGNTVLAEGPPA 330
Db 333 VNWQKSGGTRGNTVLAEGPPA 355

RESULT 14
US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 58.9%; Score 1000.5; DB 15; Length 1016;
Best Local Similarity 85.5%; Pred. No. 9.5e-68;
Matches 206; Conservative 7; Mismatches 13; Indels 15; Gaps 4;
Qy 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 67
Db 535 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 594
Qy 68 NNHVIAGATDINAFSVSGGQTYGVVDYDRTQDVAVLQLRGAGLPSAICGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVVDYDRTQDVAVLQLRGAGLPSAICGGVAVGEP 654
Qy 128 VVAMGNSGGGTPRAVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAGGPV 187
Db 655 VVAMGNSGGGTPRAVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAGGPV 714
Qy 188 VNGLGQVVGWNTAASDNFQLSGGGQFAIPICQAMAIAGQIRSGGSPTHVIGPTAFGL 238
Db 715 VNGLGQVVGWNTAASGTF--SRPG-----LPEVYLVQVSPSMGRDIKVPQSGGNNSPAVY 768
Qy 239 I 239
Db 769 L 769

RESULT 15
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-16
Query Match 58.4%; Score 992; DB 15; Length 1154;
Best Local Similarity 72.9%; Pred. No. 4.9e-67;
Matches 212; Conservative 14; Mismatches 37; Indels 28; Gaps 4;
Qy 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 67
Db 535 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 594
Qy 68 NNHVIAGATDINAFSVSGGQTYGVVDYDRTQDVAVLQLRGAGLPSAICGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVVDYDRTQDVAVLQLRGAGLPSAICGGVAVGEP 654
Qy 128 VVAMGNSGGGTPRAVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAGGPV 187
Db 655 VVAMGNSGGGTPRAVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAGGPV 714
Qy 188 VNGLGQVVGWNTAASDNFQLSGGGQFAIPICQAMAIAGQIRSGGSPTHVIGPTAFGL 247
Db 715 VNGLGQVVGWNTAASGTF-----GLLPP-----EVNSSRMVSGPGPESMLAAAAWDGV 765
Qy 248 GVVDNNGNGARVQVVVGAPASLSIGTGDVITAVDGPAPI--NSATAMADA 296
Db 766 -----AAELTSAASVSGSVVSTLIVEPWWGFAAAAAAAA 799

Search completed: June 22, 2004, 18:07:52
Job time : 30.5644 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 40.5137 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-6
Perfect score: 1698
Sequence: 1 MEHHHHAPPALPQDRFADF.....QTKSGTGTGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%
Maximum Watch 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1980s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1698	100.0	330	5 AAE29703	Aae29703 Mycobacte
2	1698	100.0	330	5 AAE17567	Aae17567 Mycobacte
3	1698	100.0	330	7 ADA26372	Ada26372 Mycobacte
4	1698	100.0	723	7 ADA26354	Ada26354 Mycobacte
5	1698	100.0	1010	7 ADA26356	Ada26356 Mycobacte
6	1695	99.8	330	5 AAE29702	Aae29702 Mycobacte
7	1695	99.8	330	5 AAE17566	Aae17566 Mycobacte
8	1688	99.4	330	7 ADA26371	Ada26371 Mycobacte
9	1642	96.7	355	2 AAY05000	Aay05000 Mycobacte
10	1642	96.7	355	4 AAO22137	Aao22137 Mycobacte
11	1642	96.7	355	4 AAG81110	Aag81110 Mycobacte
12	1642	96.7	373	2 AAY04830	Aay04830 Mycobacte
13	1631	96.1	355	2 AAW32367	Aaw32367 Mycobacte
14	1631	96.1	355	2 AAW32435	Aaw32435 Mycobacte
15	1631	96.1	355	2 AAW64307	Aaw64307 Mycobacte
16	1631	96.1	355	2 AAW81670	Aaw81670 M. tuberc
17	1631	96.1	355	2 AAY38972	Aay38972 M. tuberc
18	1631	96.1	355	2 AAY39109	Aay39109 M. tuberc
19	1631	96.1	355	4 AAU01890	Aau01890 M. tuberc
20	1631	96.1	355	5 AAE29701	Aae29701 Mycobacte
21	1631	96.1	355	5 AAE17565	Aae17565 Mycobacte
22	1000.5	58.9	1016	7 ADA26370	Ada26370 M. bovis
23	992	58.4	1154	7 ADA26368	Ada26368 Mycobacte
24	991.5	58.4	825	7 ADA26366	Ada26366 Mycobacte
25	990.5	58.3	813	7 ADA26367	Ada26367 Mycobacte

26	989	58.2	1022	7 ADA26369	Ada26369 Mycobacte
27	987	58.1	729	5 AAE29709	Aae29709 Mycobacte
28	987	58.1	729	5 AAE17573	Aae17573 Mycobacte
29	987	58.1	729	7 ADA26374	Ada26374 Mycobacte
30	984.5	58.0	875	7 ADA26365	Ada26365 Mycobacte
31	984	58.0	596	2 AAY32070	Aay32070 Mycobacte
32	984	58.0	596	5 AAE29710	Aae29710 Mycobacte
33	984	58.0	596	5 AAE17574	Aae17574 Mycobacte
34	984	58.0	599	5 AAU74599	Aau74599 Antigenic
35	984	58.0	729	4 AAO22142	Aao22142 Ra12-H9-3
36	984	58.0	729	5 AAE29708	Aae29708 Mycobacte
37	984	58.0	729	5 AAE17572	Aae17572 Mycobacte
38	984	58.0	729	7 ADA26373	Ada26373 Mycobacte
39	984	58.0	930	5 AAE29731	Aae29731 Mycobacte
40	984	58.0	930	5 ADA26364	Ada26364 Mycobacte
41	979	57.7	726	5 AAU74588	Aau74588 Antigenic
42	979	57.7	729	2 AAY32059	Aay32059 Mycobacte
43	978	57.6	195	5 AAE29704	Aae29704 Mycobacte
44	978	57.6	195	5 AAE17568	Aae17568 Mycobacte
45	676	39.8	231	2 AAY32071	Aay32071 Mycobacte

ALIGNMENTS

RESULT 1
AAE29703
ID AAE29703 standard; protein; 330 AA.
XX

AC AAE29703;

DT 27-JAN-2003 (first entry)

XX Mycobacterium tuberculosis Ra35FLMutSA mutant antigenic protein.

XX Vaccine; immunity; diagnostic agent; gene therapy; Ra35FLMutSA; mutant;
antigen; mutein.

OS Mycobacterium tuberculosis.

OS Synthetic.

FX Key Location/Qualifiers

FT Misc-difference 183

FT /note= "Wild type Ser substituted with Ala"

XX WO200272792-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US008223.

XX 13-MAR-2001; 2001US-0275837P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Brannon M, Guderian J;

XX WPI: 2002-759844/82.

XX N-PSDB; AAD47078.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
against pathogenic microorganisms e.g. Leishmania and Mycobacterium
tuberculosis.

XX Disclosure; Page 81-82; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
fusion polypeptide. The recombinant nucleic acid comprises a heterologous
polynucleotide sequence encoding an antigen or an antigenic fragment from
Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
polypeptide or its fragment. The Leishmania polynucleotide is selected
from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention

CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis Ra35FLMutSA mutant antigenic protein
 XX
 SQ Sequence 330 AA;

Query Match 100.0%; Score 1698; DB 5; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1e-123;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSDQREADFPALPLDPSAMVAQVGPVNNINIKLGYNNVAGAGTGIVID 60
 DB 1 MHHHHHAPPALSDQREADFPALPLDPSAMVAQVGPVNNINIKLGYNNVAGAGTGIVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSVGSGGTGYVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
 DB 61 PNGVLTNNHVIAGATDINAFSVGSGGTGYVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
 QY 121 GVAVGEPVVMGNSGGGTGTPRVPGRVVALGQTVQASDSLTCGAEETLNGLIQFDAAIQ 180
 DB 121 GVAVGEPVVMGNSGGGTGTPRVPGRVVALGQTVQASDSLTCGAEETLNGLIQFDAAIQ 180
 QY 181 GDAGGPVVMGNSGGGTGTPRVPGRVVALGQTVQASDSLTCGAEETLNGLIQFDAAIQ 240
 DB 181 GDAGGPVVMGNSGGGTGTPRVPGRVVALGQTVQASDSLTCGAEETLNGLIQFDAAIQ 240
 QY 241 PTAFLGLGVVDNNGNGARVQVVGSAASLGISTGCVITAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVVDNNGNGARVQVVGSAASLGISTGCVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330

RESULT 2
 AAEL17567
 ID AAE17567 standard; protein; 330 AA.
 AC AAE17567;
 XX
 XX 22-APR-2002 (first entry)
 DT
 XX Mycobacterium species Ra35FL mature protein mutant (S183A), Ra35FLMutSA.
 DE
 XX Fusion protein; antigen; serological sensitivity; immune response;
 XX tuberculosis; infection; vaccine; Ra32FLMutSA protein; mutant; mutin.
 XX Mycobacterium sp.
 OS
 XX Key Location/Qualifiers
 FH Region 8..202
 FT /note= "Ra35 N-terminal peptide"
 FT Misc-difference 183
 FT /note= "Wild type Ser substituted with Ala"
 FT Region 199..330
 FT /note= "Ra35 C-terminal peptide, Ra12"
 XX
 XX WO200198460-A2.
 XX
 XX 27-DEC-2001.
 PD
 XX 20-JUN-2001; 2001WO-US019959.
 XX
 XX 20-JUN-2000; 2000US-00597796.
 PR
 XX 01-FEB-2001; 2001US-0265737P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX

PI Skeiky Y, Reed S, Alderson M;
 XX
 DR WPI: 2002-147798/19.
 DR N-PSDB; AAD28337.
 XX
 PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 PS Claim 73; Fig 6; 136pp; English.
 XX

CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC Ra32FL mature protein mutant (S183A), Ra35FLMutSA
 XX
 SQ Sequence 330 AA;

Query Match 100.0%; Score 1698; DB 5; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1e-123;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSDQREADFPALPLDPSAMVAQVGPVNNINIKLGYNNVAGAGTGIVID 60
 DB 1 MHHHHHAPPALSDQREADFPALPLDPSAMVAQVGPVNNINIKLGYNNVAGAGTGIVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSVGSGGTGYVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
 DB 61 PNGVLTNNHVIAGATDINAFSVGSGGTGYVDVVGVDRTQDVAVLQIRGAGLPSAAIG 120
 QY 121 GVAVGEPVVMGNSGGGTGTPRVPGRVVALGQTVQASDSLTCGAEETLNGLIQFDAAIQ 180
 DB 121 GVAVGEPVVMGNSGGGTGTPRVPGRVVALGQTVQASDSLTCGAEETLNGLIQFDAAIQ 180
 QY 181 GDAGGPVVMGNSGGGTGTPRVPGRVVALGQTVQASDSLTCGAEETLNGLIQFDAAIQ 240
 DB 181 GDAGGPVVMGNSGGGTGTPRVPGRVVALGQTVQASDSLTCGAEETLNGLIQFDAAIQ 240
 QY 241 PTAFLGLGVVDNNGNGARVQVVGSAASLGISTGCVITAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVVDNNGNGARVQVVGSAASLGISTGCVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
 RESULT 3
 ADA26372
 ID ADA26372 standard; protein; 330 AA.
 XX
 XX ADA26372;
 AC
 XX 20-NOV-2003 (first entry)
 DT
 XX Mycobacterium MTB32AMutSA protein.
 DE
 XX

KW fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A; MTB85A; tuberculosis; tuberculostatic; gene therapy; vaccine; mutant; mutain.
XX Synthetic.
OS Mycobacterium sp.
XX
FH Key Location/Qualifiers
FT Misc-difference 183
FT /notes "Wild-type Ser substituted by Ala"
XX
PN WO2003070187-A2.
XX
XX
PD 28-AUG-2003.
XX
XX 18-FEB-2003; 2003WO-US004903.
XX
XX 15-FEB-2002; 2002US-0357351P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Guderian J, Reed S;
PI WPI; 2003-697554/66.
DR
DR
XX
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A and MTB39 antigens, with or without the MTB85A antigen, from a Mycobacterium species, useful for diagnosing, preventing and/or treating tuberculosis infection.
XX
XX Disclosure; Fig 19; 112pp; English.
XX
XX The invention relates to a novel nucleic acid encoding a fusion polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A polynucleotide of the invention may have a use in gene therapy, and as a vaccine. The methods and compositions of the present invention are useful for diagnosing, preventing and/or treating tuberculosis infection. The present sequence is used in the exemplification of the invention.
XX
XX
SQ Sequence 330 AA;
Query Match 100.0%; Score 1698; DB 7; Length 330;
Best Local Similarity 100.0%; Pred. No. 1e-123;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINTKLGYNNAVAGTGIVID 60
Db 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINTKLGYNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQLRGAGLPSAAIIG 120
Db 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQLRGAGLPSAAIIG 120
QY 121 GVAVGEPVVMGNSGGGQTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIOP 180
Db 121 GVAVGEPVVMGNSGGGQTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIOP 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQSLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVNVGLGVVGMNTAASDNFQSLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
RESULT 4
AD26354

ID ADA26354 standard; protein; 723 AA.
XX ADA26354;
XX 20-NOV-2003 (first entry)
DT
XX Mycobacterium MTB32-MTB39F fusion protein.
DE
XX fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A; tuberculosis; tuberculostatic; gene therapy; vaccine.
KW
XX Chimeric.
OS Mycobacterium sp.
XX WO2003070187-A2.
XX
XX 28-AUG-2003.
XX
XX 18-FEB-2003; 2003WO-US004903.
XX
XX 15-FEB-2002; 2002US-0357351P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Guderian J, Reed S;
PI WPI; 2003-697554/66.
XX
DR N-PSDB; ADA26353.
DR
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A and MTB39 antigens, with or without the MTB85A antigen, from a Mycobacterium species, useful for diagnosing, preventing and/or treating tuberculosis infection.
XX
XX Claim 4; Fig 2; 112pp; English.
XX
XX The invention relates to a novel nucleic acid encoding a fusion polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A polynucleotide of the invention may have a use in gene therapy, and as a vaccine. The methods and compositions of the present invention are useful for diagnosing, preventing and/or treating tuberculosis infection. The present sequence is used in the exemplification of the invention.
XX
XX
SQ Sequence 723 AA;
Query Match 100.0%; Score 1698; DB 7; Length 723;
Best Local Similarity 100.0%; Pred. No. 2.7e-123;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINTKLGYNNAVAGTGIVID 60
Db 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVQVNNINTKLGYNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQLRGAGLPSAAIIG 120
Db 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQLRGAGLPSAAIIG 120
QY 121 GVAVGEPVVMGNSGGGQTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIOP 180
Db 121 GVAVGEPVVMGNSGGGQTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIOP 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQSLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVNVGLGVVGMNTAASDNFQSLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

```

Db      301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
RESULT 5
ADA26356
ID      ADA26356 standard; protein; 1010 AA.
XX
AC      ADA26356;
XX
DT      20-NOV-2003 (first entry)
XX
DE      Mycobacterium MTB-102F fusion protein.
XX
KW      fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW      tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
OS      Chimeric.
OS      Mycobacterium sp.
XX
FN      WO2003070187-A2.
XX
PD      28-AUG-2003.
XX
PF      18-FEB-2003; 2003WO-US0004903.
XX
PR      15-FEB-2002; 2002US-0357351P.
XX
PA      (CORI-) CORIXA CORP.
XX
PI      Skeiky Y, Guderian J, Reed S;
XX
DR      WPI; 2003-697554/66.
DR      N-PSDB; ADA26355.
XX
XX      New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT      and MTB39 antigens, with or without the MTB85A antigen, from a
PT      Mycobacterium species, useful for diagnosing, preventing and/or treating
PT      tuberculosis infection.
XX
PS      Claim 6; Fig 4; 112pp; English.
XX
XX      The invention relates to a novel nucleic acid encoding a fusion
CC      polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC      MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC      A polypeptide of the invention has tuberculostatic activity. A
CC      polynucleotide of the invention may have a use in gene therapy, and as a
CC      vaccine. The methods and compositions of the present invention are useful
CC      for diagnosing, preventing and/or treating tuberculosis infection. The
CC      present sequence is used in the exemplification of the invention.
XX
SQ      Sequence 1010 AA;

Query Match      100.0%; Score 1698; DB 7; Length 1010;
Best Local Similarity 100.0%; Pred. No. 4,1e-123;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNNAVAGTGIVID 60
Db      1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNNAVAGTGIVID 60

Qy      61 PNGVLTNNHVIAGATDINAFSVGSGGTGYDVVGVYDRTQDVAVLQIRAGGLPSAAIGG 120
Db      61 PNGVLTNNHVIAGATDINAFSVGSGGTGYDVVGVYDRTQDVAVLQIRAGGLPSAAIGG 120

Qy      121 GVAVGEPPVAMGNSGGGGTPRAVGRVVALGQTVQASDLSLTGAETLNLQIFDAAIQP 180
Db      121 GVAVGEPPVAMGNSGGGGTPRAVGRVVALGQTVQASDLSLTGAETLNLQIFDAAIQP 180

Qy      181 GDAGGPPVNLGQVGVGMNTAASNDFOLSQGGQFAIPIGQAMAIAQGIIRSGGSPFVHIG 240
Db      181 GDAGGPPVNLGQVGVGMNTAASNDFOLSQGGQFAIPIGQAMAIAQGIIRSGGSPFVHIG 240

Qy      241 PTFALGLGVDDNNGNGARVQRVWGSAPASLIGTSDVITAVDGPAINSATAMADALNGH 300

```

QY 61 PNGVLTNNHVIAGATDINAFSVSGQTYGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
 DB 61 PNGVLTNNHVIAGATDINAFSVSGQTYGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVAMGNSGGGGTPRAVGRVWALGQTVOASDSLGTGAEETLNGLIQFDAAIQP 180
 DB 121 GVAVGEPVAMGNSGGGGTPRAVGRVWALGQTVOASDSLGTGAEETLNGLIQFDAAIQP 180
 QY 181 GDAGGPVNGLGQVVGNTAAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 DB 181 GDSGGPVNGLGQVVGNTAAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTFALGLGVVDNNGGARVORVVGSAAPASISLGISTGDIVITAVDGAPINSATADALNGH 300
 DB 241 PTFALGLGVVDNNGGARVORVVGSAAPASISLGISTGDIVITAVDGAPINSATADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 7

AAE17566
 ID AAE17566 standard; protein; 330 AA.
 AC AAE17566;

22-APR-2002 (first entry)

Mycobacterium sp. MTB32A (Ra35FL) mature protein.

XX Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
 XX Mycobacterium sp.

XX Key Location/Qualifiers
 FH Region 8..202
 FT /note= "Ra35 N-terminal peptide"
 FT Misc-difference 182
 FT /note= "Encoded by GAG"
 FT Misc-difference 183
 FT /note= "Encoded by GCG"
 FT Region 199..330
 FT /note= "Ra35 C-terminal peptide, Ra12"
 XX
 XX WO2001198460-A2.

27-DEC-2001.

20-JUN-2001; 2001WO-US019959.

20-JUN-2000; 2000US-00597796.

01-FEB-2001; 2001US-0265737P.

(CORI-) CORIXA CORP.

Skeiky Y, Reed S, Alderson M;

WPI; 2002-147798/19.

N-PSDB; AAD28336.

XX Composition comprising MTB39 antigen and MTB32A antigen from
 FT Mycobacterium species, useful for eliciting immune response in a subject.
 XX

PS Claim 70; Fig 6; 136pp; English.

CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase

CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) mature protein
 XX
 SQ Sequence 330 AA;

Query Match 99.8%; Score 1695; DB 5; Length 330;
 Best Local Similarity 99.7%; Fred. No. 1.7e-123;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVVNINIKLGYNNVAGAGTGIVID 60
 DB 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQVGPVVNINIKLGYNNVAGAGTGIVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSVSGQTYGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
 DB 61 PNGVLTNNHVIAGATDINAFSVSGQTYGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVAMGNSGGGGTPRAVGRVWALGQTVOASDSLGTGAEETLNGLIQFDAAIQP 180
 DB 121 GVAVGEPVAMGNSGGGGTPRAVGRVWALGQTVOASDSLGTGAEETLNGLIQFDAAIQP 180
 QY 181 GDAGGPVNGLGQVVGNTAAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 DB 181 GDSGGPVNGLGQVVGNTAAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTFALGLGVVDNNGGARVORVVGSAAPASISLGISTGDIVITAVDGAPINSATADALNGH 300
 DB 241 PTFALGLGVVDNNGGARVORVVGSAAPASISLGISTGDIVITAVDGAPINSATADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 8

ADA26371
 ID ADA26371 standard; protein; 330 AA.
 XX

AC ADA26371;

DT 20-NOV-2003 (first entry)

DE Mycobacterium wild type MTB32A protein.

XX fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 XX Chimeric.

OS Mycobacterium sp.

PN WO2003070187-A2.

PD 28-AUG-2003.

XX 18-FEB-2003; 2003WO-US004903.

XX 15-FEB-2002; 2002US-0357351P.

PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Guderian J, Reed S;
 XX DR WPI; 2003-697554/66.
 XX PI New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 XX PT and MTB39 antigens, with or without the MTB85A antigen, from a
 XX PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 XX PT tuberculosis infection.
 XX PS Disclosure; Fig 19; 112pp; English.
 XX CC The invention relates to a novel nucleic acid encoding a fusion
 XX CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 XX CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 XX CC A polypeptide of the invention has tuberculostatic activity. A
 XX CC polynucleotide of the invention may have a use in gene therapy, and as a
 XX CC vaccine. The methods and compositions of the present invention are useful
 XX CC for diagnosing, preventing and/or treating tuberculosis infection. The
 XX CC present sequence is used in the exemplification of the invention.
 XX CC
 XX CC Sequence 330 AA;
 XX CC
 XX CC Query Match 99.4%; Score 1688; DB 7; Length 330;
 XX CC Best Local Similarity 99.4%; Pred. No. 6.1e-123;
 XX CC Matches 328; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGTGVID 60
 DB 1 MHHHHHAPPALSDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGTGVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVVGVDRTQDVAVLQLRGAGGLPSAAIG 120
 DB 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVVGVDRTQDVAVLQLRGAGGLPSAAIG 120
 QY 121 GVAVGEPVWAMNSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
 DB 121 GVAVGEPVWAMNSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
 QY 181 GDAGGPPVNLGQVVGMMNTAASDNFQLSQGGQGFPAIPIGQAMAIAGQIRSGGSPVTHIG 240
 DB 181 GDAGGPPVNLGQVVGMMNTAASDNFQLSQGGQGFPAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVNDNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVNDNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 RESULT 9
 ID AAY05000 standard; protein; 355 AA.
 XX AC AAY05000;
 XX DT 06-JUL-1999 (first entry)
 XX DE Mycobacterium species protein sequence 50D.
 XX KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 XX KW hybridisation; detection; vaccine; immunisation; infection.
 XX KW Mycobacterium sp.
 XX OS WO9909186-A2.
 XX PN 25-FEB-1999.
 XX PD 14-AUG-1998; 98WO-FR001813.
 XX PF 14-AUG-1998; 98WO-FR001813.
 XX XX

PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX PA (INSP) INST PASTEUR.
 XX PI Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 XX PI Goguet De La Salmoniere Y;
 XX DR WPI; 1999-181045/15.
 XX DR N-PSDB; AAX34251.
 XX PT Mycobacterial DNA vectors containing reporter constructs - for
 XX PT identifying coding or promoter sequences involved in infection-associated
 XX PT protein expression.
 XX PS Claim 32; Fig 50D; 309pp; French.
 XX CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 XX CC from various Mycobacterium species microorganisms. The encoding
 XX CC nucleotide sequences can be used as primers and probes for methods for
 XX CC detecting and identifying mycobacteria, especially belonging to the M.
 XX CC tuberculosis complex. The encoded proteins can be used in vaccines for
 XX CC immunisation against a bacterial or viral infection
 XX CC
 XX CC Sequence 355 AA;
 XX CC
 XX CC Query Match 96.7%; Score 1642; DB 2; Length 355;
 XX CC Best Local Similarity 99.7%; Pred. No. 2.5e-119;
 XX CC Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGTGVIDPENGWVLT 67
 DB 33 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGTGVIDPENGWVLT 92
 QY 68 NNHVIAGATDINAFSVSGGQTYGVVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEP 127
 DB 93 NNHVIAGATDINAFSVSGGQTYGVVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEP 152
 QY 128 VVAMNSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 187
 DB 153 VVAMNSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 212
 QY 188 NVNGLGQVVGMMNTAASDNFQLSQGGQGFPAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 247
 DB 213 NVNGLGQVVGMMNTAASDNFQLSQGGQGFPAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 272
 QY 248 GVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
 DB 273 GVVDNNGNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355
 RESULT 10
 ID AAO22137
 XX AC AAO22137 standard; protein; 355 AA.
 XX DT 03-OCT-2002 (first entry)
 XX DE Mycobacterium tuberculosis MTB32A protein.
 XX KW Mycobacterium tuberculosis MTB32A protein.
 XX KW Ral2; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
 XX KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
 XX KW immunogen; cytokine.
 XX OS Mycobacterium tuberculosis.
 XX PN WO200125401-A2.
 XX XX

PD 12-APR-2001.
XX
XX
PF 06-OCT-2000; 2000WO-US027652.
XX
XX
PR 07-OCT-1999; 99US-0158585P.
XX
XX
PA (CORI-) CORIXA CORP.
XX
XX
PI Skeiky Y, Guderian J;
XX
XX
DR WPI; 2001-266299/27.
DR N-PSDB; AAL40768.
XX
XX
XX
PT Recombinant nucleic acid molecule for producing high yield expression of
PT desired fusion polypeptides, encodes fusion polypeptide comprising
PT Mycobacterium tuberculosis coding sequence and heterologous polypeptide.
XX
XX
PS Disclosure; Fig 1; 39pp; English.
XX
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide, comprising a polynucleotide sequence of Ral2, a 14
CC kDa C-terminal fragment of serine protease antigen MTB32A of
CC Mycobacterium tuberculosis, and a heterologous polynucleotide sequence.
CC The recombinant fusion nucleic acids and polypeptides are useful for
CC providing stable and high yield expression of fusion polypeptides of both
CC eukaryotic and prokaryotic origin and to encode a protein product for use
CC as an antigen for detecting serum antibodies. The presence of serum
CC antibodies to M. tuberculosis antigens in an individual indicates that
CC the individual is infected with it. The fusion polypeptides are useful as
CC sources of proteins for monitoring binding of serum antibodies to fusion
CC proteins and as an immunogen to induce and/or enhance immune responses.
CC The coding sequences can be ligated with a coding sequence of another
CC molecule such as M. tuberculosis antigens, a cytokine or an adjuvant and
CC can be used in vivo as a DNA vaccine. This sequence represents the
CC Mycobacterium tuberculosis MTB32A protein
XX
SQ Sequence 355 AA;

Query Match 96.7%; Score 1642; DB 4; Length 355;
Best Local Similarity 99.7%; Pred. No. 2.5e-119;
Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGVIDPENGVLVT 67
Db 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGVIDPENGVLVT 92

QY 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDYDRTQDVAVLQLRGAGLPSAIGGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVDVVGVDYDRTQDVAVLQLRGAGLPSAIGGGVAVGEP 152

QY 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPV 212

QY 188 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPV 247
Db 213 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPV 272

QY 248 GVVDNNGNGARVQRVVGSAAPASLGTSTGDTVITAVDVGAPINSATAMADALNGHHPGDVIS 307
Db 273 GVVDNNGNGARVQRVVGSAAPASLGTSTGDTVITAVDVGAPINSATAMADALNGHHPGDVIS 332

QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VTWQTKSGGTRTGNVTLAEGPPA 355

RESULT 11
AAG81110
ID AAG81110 standard; protein; 355 AA.
XX
AC AAG81110;
XX

DT 04-SEP-2001 (first entry)
XX
XX
DE Mycobacterium tuberculosis potential drug target protein SEQ ID 161.
XX
XX
KW Drug target; growth; organism viability; characterisation.
XX
XX
OS Mycobacterium tuberculosis.
XX
XX
PN WO200135317-A1.
XX
XX
PD 17-MAY-2001.
XX
XX
PF 13-NOV-2000; 2000WO-US031152.
XX
XX
PR 12-NOV-1999; 99US-0165086P.
PR 12-NOV-1999; 99US-0165124P.
PR 01-FEB-2000; 2000US-0179531P.
XX
XX
PA (REGC) UNIV CALIFORNIA.
XX
XX
PI Eisenberg D, Rotstein SH, Marcotte EM;
XX
XX
DR WPI; 2001-329193/34.
DR N-PSDB; AAH51961.
XX
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the sequences.
XX
XX
PS Disclosure; Page 157; 207pp; English.
XX
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism
XX
SQ Sequence 355 AA;

Query Match 96.7%; Score 1642; DB 4; Length 355;
Best Local Similarity 99.7%; Pred. No. 2.5e-119;
Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGVIDPENGVLVT 67
Db 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGVIDPENGVLVT 92

QY 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDYDRTQDVAVLQLRGAGLPSAIGGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVDVVGVDYDRTQDVAVLQLRGAGLPSAIGGGVAVGEP 152

QY 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAGGPV 187
Db 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAGGPV 212

QY 188 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAGGPV 247
Db 213 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAGGPV 272

QY 248 GVVDNNGNGARVQRVVGSAAPASLGTSTGDTVITAVDVGAPINSATAMADALNGHHPGDVIS 307
Db 273 GVVDNNGNGARVQRVVGSAAPASLGTSTGDTVITAVDVGAPINSATAMADALNGHHPGDVIS 332

QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VTWQTKSGGTRTGNVTLAEGPPA 355

Db 333 VTWTKSGGTRTGNVTLAEGPPA 355

RESULT 12

AA04830
ID AAY04830 standard; protein; 379 AA.

XX AC AAY04830;

XX DT 06-JUL-1999 (first entry)

XX XX Mycobacterium species protein sequence 50F.

XX XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX KW hybridisation; detection; vaccine; immunisation; infection.

XX XX Mycobacterium sp.

XX FN WO9909186-A2.

XX PD 25-FEB-1999.

XX PD 14-AUG-1998; 98WO-FR001813.

XX PR 14-AUG-1997; 97FR-00010404.

XX PR 11-SEP-1997; 97FR-00011325.

XX XX (INSP) INST PASTEUR.

XX PI Gicquel B, Portnoie D, Lim B, Pelicic V, Guigueno A;

XX PI Goguet De La Salmoniere Y;

XX DR WPI; 1999-181045/15.

XX DR N-PSDB; AAX34252.

XX PT Mycobacterial DNA vectors containing reporter constructs - for
XX PT identifying coding or promoter sequences involved in infection-associated
XX PT protein expression.

XX PS Claim 32; Fig 50F; 309pp; French.

XX XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
XX CC from various Mycobacterium species microorganisms. The encoding
XX CC nucleotide sequences can be used as primers and probes for methods for
XX CC detecting and identifying mycobacteria, especially belonging to the M.
XX CC tuberculosis complex. The encoded proteins can be used in vaccines for
XX CC immunisation against a bacterial or viral infection

XX SQ Sequence 379 AA;

Query Match 96.7%; Score 1642; DB 2; Length 379;

Best Local Similarity 99.7%; Pred. No. 2.7e-119; Indels 0; Gaps 0;
Matches 322; Conservative 1; Mismatches 0;

Qy 8 APPALSQDRFADFPALPLDPSAMVAQVQVNTKLGYNNAVAGTGIVDPNGVLT 67

Db 57 APPALSQDRFADFPALPLDPSAMVAQVQVNTKLGYNNAVAGTGIVDPNGVLT 116

Qy 68 NNHVIAGATDINAFSVGSGQTYGVDDVGYDRTQDVAVLQLRGAGLPSAALGGGVAVGEP 127

Db 117 NNHVIAGATDINAFSVGSGQTYGVDDVGYDRTQDVAVLQLRGAGLPSAALGGGVAVGEP 176

Qy 128 VVAMGNSGGQGTTPRAVPGRVVAGLQGVQASDSLTGAETLNGLIQFDAAIQPDAGGPV 187

Db 177 VVAMGNSGGQGTTPRAVPGRVVAGLQGVQASDSLTGAETLNGLIQFDAAIQPDAGGPV 236

Qy 188 VNLGQVGVGMNTAASDNFQLSQGGQGAIPICQMAIAGQIRSGGSPVTHIGTAFGL 247

Db 237 VNLGQVGVGMNTAASDNFQLSQGGQGAIPICQMAIAGQIRSGGSPVTHIGTAFGL 296

Qy 248 GVVDNNGNGARVQVWGSAASLGISTGVITAVDGAIPNSATAMADALNGHHPGDVIS 307

Db 297 GVVDNNGNGARVQVWGSAASLGISTGVITAVDGAIPNSATAMADALNGHHPGDVIS 356

Qy 308 VTWTKSGGTRTGNVTLAEGPPA 330

Db 357 VTWTKSGGTRTGNVTLAEGPPA 379

RESULT 13

AAW32367

ID AAW32367 standard; protein; 355 AA.

XX AC AAW32367;

XX DT 13-JAN-1998 (first entry)

XX XX Mycobacterium tuberculosis antigen TbrA35.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX KW skin testing; M.tuberculosis.

XX OS Mycobacterium tuberculosis.

XX PN WO9709429-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US014675.

XX PR 01-SEP-1995; 95US-00523435.

XX PR 22-SEP-1995; 95US-00532136.

XX PR 22-MAR-1996; 96US-00620280.

XX PR 05-JUN-1996; 96US-00658800.

XX PR 12-JUL-1996; 96US-00680573.

XX PA (CORI-) CORIXA CORP.

XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedwick TH, Twardzik DR;

XX DR WPI; 1997-192904/17.

XX DR N-PSDB; AAT91414.

XX PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
XX PT useful for diagnosis of M. tuberculosis infection.

XX PS Example 3; Page 124-126; 190pp; English.

XX XX A new immunogenic polypeptide has been developed comprising an
XX CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
XX CC variant differing only in conservative substitutions and/or
XX CC modifications). The present sequence represents a M.tuberculosis antigen,
XX CC TbrA35. The immunogenic polypeptide can be used to diagnose
XX CC M.tuberculosis infection by forming complexes with specific antibodies in
XX CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be
XX CC used as diagnostic primers or probes and agents that bind to the antigen,
XX CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
XX CC also used for diagnosis

XX SQ Sequence 355 AA;

Query Match 96.1%; Score 1631; DB 2; Length 355;

Best Local Similarity 99.1%; Pred. No. 1.8e-118;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 8 APPALSQDRFADFPALPLDPSAMVAQVQVNTKLGYNNAVAGTGIVDPNGVLT 67

Db 33 APPALSQDRFADFPALPLDPSAMVAQVQVNTKLGYNNAVAGTGIVDPNGVLT 92

Qy 68 NNHVIAGATDINAFSVGSGQTYGVDDVGYDRTQDVAVLQLRGAGLPSAALGGGVAVGEP 127

Db 93 NNHVIAGATDINAFSVGSGQTYGVDDVGYDRTQDVAVLQLRGAGLPSAALGGGVAVGEP 152

Qy 128 VVAMGNSGGQGTTPRAVPGRVVAGLQGVQASDSLTGAETLNGLIQFDAAIQPDAGGPV 187

Db 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGPV 212
 QY 188 VNLGQVVGMNTAASDNFQLSQGGGFAIPGOAMAIAGQIRSGGSPVHIGTFLGL 247
 Db 213 VNLGQVVGMNTAASDNFQLSQGGGFAIPGOAMAIAGQIRSGGSPVHIGTFLGL 272
 QY 248 GVVDNNGGARVORVVGSAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 307
 Db 273 GVVDNNGGARVORVVGSAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
 Db 333 VNWQKSGGTRTGNVTLAEGPPA 355

 RESULT 14
 AAW32435
 ID AAW32435 standard; protein; 355 AA.
 XX
 AC AAW32435;
 XX
 DT 08-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TbrA35.
 XX
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9709428-A2.
 XX
 PD 13-MAR-1997.
 XX
 PF 30-AUG-1996; 96WO-US014674.
 XX
 PR 01-SEP-1995; 95US-00523436.
 PR 22-SEP-1995; 95US-00533634.
 PR 22-MAR-1996; 96US-00620874.
 PR 05-JUN-1996; 96US-00659683.
 PR 12-JUL-1996; 96US-00680574.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TH, Twardzik DR;
 XX
 XX WPI; 1997-192903/17.
 DR N-PSDB; AAT91477.
 XX
 XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.
 XX
 XX Example 3; Page 114-116; 168pp; English.
 PS
 XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbrA35. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)
 XX
 XX Sequence 355 AA;
 SS
 Query Match 96.1%; Score 1631; DB 2; Length 355;
 Best Local Similarity 99.1%; Pred. No. 1.8e-118;
 Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

 QY 8 APPALSQDRFADFPALPLDPSAMVAQVQVNVNITKLGYNNAVAGTIVIDPNGVLT 67

Db 33 APPALSQDRFADFPALPLDPSAMVAQVQVNVNITKLGYNNAVAGTIVIDPNGVLT 92
 QY 68 NNHVIAGATDINAFVSGGQTYGVVDVGYDRITQDVAVIQLRGAGGLPSAAICGGVAVGEP 127
 Db 93 NNHVIAGATDINAFVSGGQTYGVVDVGYDRITQDVAVIQLRGAGGLPSAAICGGVAVGEP 152
 QY 128 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGPV 187
 Db 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGPV 212
 QY 188 VNLGQVVGMNTAASDNFQLSQGGGFAIPGOAMAIAGQIRSGGSPVHIGTFLGL 247
 Db 213 VNLGQVVGMNTAASDNFQLSQGGGFAIPGOAMAIAGQIRSGGSPVHIGTFLGL 272
 QY 248 GVVDNNGGARVORVVGSAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 307
 Db 273 GVVDNNGGARVORVVGSAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
 Db 333 VNWQKSGGTRTGNVTLAEGPPA 355

 RESULT 15
 AAW64307
 ID AAW64307 standard; protein; 355 AA.
 XX
 AC AAW64307;
 XX
 DT 17-OCT-2003 (revised)
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TbrA35.
 XX
 KW Tuberculosis; infection; diagnosis; antigen; TbrA35.
 XX
 OS Mycobacterium tuberculosis; strain H37Ra.
 XX
 PN WO9816645-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018214.
 XX
 PR 11-OCT-1996; 96US-00729622.
 PR 13-MAR-1997; 97US-00818111.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 XX WPI; 1998-251292/22.
 DR N-PSDB; AAW44355.
 XX
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 PS Example 3; Page 115-116; 250pp; English.
 XX
 CC This polypeptide comprises Mycobacterium tuberculosis soluble antigen
 CC TbrA35. It is encoded by a DNA sequence (see AAW44355) isolated from a M.
 CC tuberculosis strain H37Ra expression library with rabbit anti-sera raised
 CC against M. tuberculosis supernatant. No significant homology was found
 CC between TbrA35 and Genbank database sequences. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64373) comprising an antigenic portion of a
 CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic kits for

CC detecting M. tuberculosis infection in a patient using the above
CC polypeptides, antibodies or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
CC field)
XX
SQ Sequence 355 AA;

Query Match	96.1%;	Score 1631;	DB 2;	Length 355;
Best Local Similarity	99.1%;	Pred. No. 1.8e-118;		
Matches 320;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	8	APPALSDRFPALPLDPSAMVAQVGVNINIKLGYNNAVGAGTGIVIDPNGWILT	67
Db	33	APPALSDRFPALPLDPSAMVAQVGVNINIKLGYNNAVGAGTGIVIDPNGWILT	92
Qy	68	NNHVIAGATDINAFSVGSGGTGYVDVVGYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP	127
Db	93	NNHVIAGATDINAFSVGSGGTGYVDVVGYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP	152
Qy	128	VWAMNSGGGGTTPRAYGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV	187
Db	153	VWAMNSGGGGTTPRAYGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV	212
Qy	188	VNGLGVVGMNTAASDNFQSQGGGFAPIGQAMAIAGQIRSGGSPVHIGTAFGL	247
Db	213	VNGLGVVGMNTAASDNFQSQGGGFAPIGQAMAIAGQIRSGGSPVHIGTAFGL	272
Qy	248	GVVDNNGNGARVQVVGSAAPASLIGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS	307
Db	273	GVVDNNGNGARVQVVGSAAPASLIGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS	332
Qy	308	VTWTKSGGTGTGNVTLAEGPPA	330
Db	333	VNWTKSGGTGTGNVTLAEGPPA	355

Search completed: June 22, 2004, 17:15:25
Job time : 41.5137 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 11.78 Seconds

(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-6

Perfect score: 1698

Sequence: 1 MHHHHHAPPALSQDRFADF.....QTKSGGTGTGNTLAEGPPA 330

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*

2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*

3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*

4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*

5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*

6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1631	96.1	355	3	US-08-818-112-79
2	1631	96.1	355	4	US-08-818-111-80
3	1631	96.1	355	4	US-09-556-556-79
4	1631	96.1	355	4	US-09-072-598-80
5	1631	96.1	355	4	US-09-072-967-79
6	984	58.0	596	4	US-09-287-849-26
7	984	58.0	729	4	US-09-223-040-2
8	984	58.0	729	4	US-09-287-849-2
9	676	39.8	231	4	US-09-287-849-28
10	670	39.5	132	3	US-08-818-112-66
11	670	39.5	132	4	US-08-818-111-67
12	670	39.5	132	4	US-09-556-556-66
13	670	39.5	132	4	US-09-072-596-67
14	670	39.5	132	4	US-09-072-967-66
15	670	39.5	132	4	US-09-636-215-819
16	670	39.5	132	4	US-09-636-215-848
17	670	39.5	132	4	US-09-585-166A-819
18	670	39.5	132	4	US-09-585-166A-848
19	633	37.3	224	4	US-09-636-215-825
20	633	37.3	224	4	US-09-636-215-825
21	633	37.3	267	4	US-09-643-597-352
22	633	37.3	267	4	US-09-606-421B-352
23	633	37.3	273	4	US-09-736-457-1864
24	633	37.3	299	4	US-09-643-597-354
25	633	37.3	299	4	US-09-606-421B-354
26	633	37.3	304	4	US-09-636-215-835
27	633	37.3	304	4	US-09-685-166A-835

28	633	37.3	314	4	US-09-736-457-1863	Sequence 1863, App
29	633	37.3	400	4	US-09-636-215-852	Sequence 852, App
30	633	37.3	400	4	US-09-685-166A-852	Sequence 852, App
31	633	37.3	487	4	US-09-620-412C-349	Sequence 349, App
32	633	37.3	487	4	US-09-598-419-349	Sequence 349, App
33	633	37.3	518	4	US-09-620-412C-333	Sequence 333, App
34	633	37.3	518	4	US-09-598-419-333	Sequence 333, App
35	633	37.3	525	4	US-09-556-877-196	Sequence 196, App
36	633	37.3	525	4	US-09-620-412C-196	Sequence 196, App
37	633	37.3	525	4	US-09-598-419-196	Sequence 196, App
38	633	37.3	583	4	US-09-620-412C-353	Sequence 353, App
39	633	37.3	583	4	US-09-598-419-353	Sequence 353, App
40	633	37.3	585	4	US-09-620-412C-337	Sequence 337, App
41	633	37.3	585	4	US-09-598-419-337	Sequence 337, App
42	633	37.3	619	4	US-09-620-412C-309	Sequence 309, App
43	633	37.3	619	4	US-09-598-419-309	Sequence 309, App
44	633	37.3	631	4	US-09-620-412C-325	Sequence 325, App
45	633	37.3	631	4	US-09-598-419-325	Sequence 325, App

ALIGNMENTS

RESULT 1
US-08-818-112-79
; Sequence 79, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/POCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-79

Query Match 96.1%; Score 1631; DB 3; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGAGTGIVDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNTKLGYNNAVAGAGTGIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNSGGGGTTPRAVGRVVALQVQASDSLTGAETLNGLIQFDDAAIQDSDGPGV 187
DB 153 VVAMNSGGGGTTPRAVGRVVALQVQASDSLTGAETLNGLIQFDDAAIQDSDGPGV 212
QY 188 VVNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFLGL 247
DB 213 VVNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFLGL 272
QY 248 GVVDNNGGARVQRVVGSAAPASISLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGGARVQRVVGSAAPASISLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2
US-08-818-111-80
; Sequence 80, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-80
Query Match 96.1%; Score 1631; DB 4; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGAGTGIVDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNTKLGYNNAVAGAGTGIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNSGGGGTTPRAVGRVVALQVQASDSLTGAETLNGLIQFDDAAIQDSDGPGV 187
DB 153 VVAMNSGGGGTTPRAVGRVVALQVQASDSLTGAETLNGLIQFDDAAIQDSDGPGV 212
QY 188 VVNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFLGL 247
DB 213 VVNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFLGL 272
QY 248 GVVDNNGGARVQRVVGSAAPASISLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGGARVQRVVGSAAPASISLGISTGVDVITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3
US-09-056-556-79
; Sequence 79, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79
Query Match 96.1%; Score 1631; DB 4; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGAGTGIVDPNGVLT 67

Db 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINTKLGYNNNAVAGTGIVIDPNGVVL 92
QY 68 NNHVIAGATDINAFSVGSGGTQYVDVVGDRDQDVAVLQLRGAGGLPSAAICGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVGSGGTQYVDVVGDRDQDVAVLQLRGAGGLPSAAICGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPRVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDAGGPV 187
Db 153 VVAMGNSGGGGTTPRAVPRVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDAGGPV 212
QY 188 VVAMGNSGGGGTTPRAVPRVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDAGGPV 247
Db 213 VVAMGNSGGGGTTPRAVPRVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDAGGPV 272
QY 248 GVVDNNGNGARVQVVGSAAPASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVIS 307
Db 273 GVVDNNGNGARVQVVGSAAPASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4
US-09-072-596-80
; Sequence 80, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-80

Query Match 96.1%; Score 1631; DB 4; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINTKLGYNNNAVAGTGIVIDPNGVVL 67
Db 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINTKLGYNNNAVAGTGIVIDPNGVVL 92
QY 68 NNHVIAGATDINAFSVGSGGTQYVDVVGDRDQDVAVLQLRGAGGLPSAAICGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVGSGGTQYVDVVGDRDQDVAVLQLRGAGGLPSAAICGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPRVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDAGGPV 187
Db 153 VVAMGNSGGGGTTPRAVPRVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDAGGPV 212
QY 188 VVAMGNSGGGGTTPRAVPRVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDAGGPV 247
Db 213 VVAMGNSGGGGTTPRAVPRVVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDAGGPV 272
QY 248 GVVDNNGNGARVQVVGSAAPASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVIS 307
Db 273 GVVDNNGNGARVQVVGSAAPASLSLTGSDVITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 5
US-09-072-967-79
; Sequence 79, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-79

Query Match 96.1%; Score 1631; DB 4; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVGAGTGIVDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVGAGTGIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPDAGGPV 187
DB 153 VVAMNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPDAGGPV 212
QY 188 VVAMNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPDAGGPV 247
DB 213 VVAMNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPDAGGPV 272
QY 248 GVVNNNGNGARVQVVGSAFPAASLGISTGDTITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVNNNGNGARVQVVGSAFPAASLGISTGDTITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VVWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VVWQTKSGGTRTGNVTLAEGPPA 355

RESULT 6

US-09-287-849-26

; Sequence 26, Application US/09287849

; Patent No. 6627198

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Alderson, Mark

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens

; FILE REFERENCE: 014058-009020US

; CURRENT APPLICATION NUMBER: US/09/287,849

; CURRENT FILING DATE: 1999-04-07

; PRIOR APPLICATION NUMBER: US 08/818,112

; PRIOR FILING DATE: 1997-03-13

; PRIOR APPLICATION NUMBER: US 08/942,578

; PRIOR FILING DATE: 1997-10-01

; PRIOR APPLICATION NUMBER: US 09/025,197

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 09/056,556

; PRIOR FILING DATE: 1998-04-07

; PRIOR APPLICATION NUMBER: US 09/223,040

; PRIOR FILING DATE: 1998-12-30

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 26

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion

US-09-287-849-26

Query Match 58.0%; Score 984; DB 4; Length 596;

Best Local Similarity 99.5%; Pred. No. 1.6e-77;

Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVGAGTGIVDPNGVLT 67
DB 402 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVGAGTGIVDPNGVLT 461

QY 68 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 462 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 521
QY 128 VVAMNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPDAGGPV 187
DB 522 VVAMNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPDAGGPV 581
QY 188 VVGLGQVVGMNTAAS 202
DB 582 VVGLGQVVGMNTAAS 596

RESULT 7

US-09-223-040-2

; Sequence 2, Application US/09223040

; Patent No. 6544522

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Alderson, Mark

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens

; FILE REFERENCE: 014058-009010US

; CURRENT APPLICATION NUMBER: US/09/223,040

; CURRENT FILING DATE: 1998-12-30

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 729

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion

US-09-223-040-2

Query Match 58.0%; Score 984; DB 4; Length 729;

Best Local Similarity 99.5%; Pred. No. 2.2e-77;

Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVGAGTGIVDPNGVLT 67
DB 535 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVGAGTGIVDPNGVLT 594
QY 68 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 595 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 654
QY 128 VVAMNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPDAGGPV 187
DB 655 VVAMNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPDAGGPV 714
QY 188 VVGLGQVVGMNTAAS 202
DB 715 VVGLGQVVGMNTAAS 729

RESULT 8

US-09-287-849-2

; Sequence 2, Application US/09287849

; Patent No. 6627198

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Alderson, Mark

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens

; FILE REFERENCE: 014058-009020US

```
/ CURRENT APPLICATION NUMBER: US/09/287,849
/ CURRENT FILING DATE: 1999-04-07
/ PRIOR APPLICATION NUMBER: US 08/818,112
/ PRIOR FILING DATE: 1997-03-13
/ PRIOR APPLICATION NUMBER: US 08/942,578
/ PRIOR FILING DATE: 1997-10-01
/ PRIOR APPLICATION NUMBER: US 09/025,197
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 09/056,556
/ PRIOR FILING DATE: 1998-04-07
/ PRIOR APPLICATION NUMBER: US 09/223,040
/ PRIOR FILING DATE: 1998-12-30
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 729
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

Query Match          58.0%; Score 984; DB 4; Length 729;
Best Local Similarity 99.5%; Pred. No. 2.2e-77;
Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 APPALSQDRFADFPALPLDPSAMVAGVGVVNNINTKLGYNNVAGAGTGIVIDPNGVLT 67
Db      535 APPALSQDRFADFPALPLDPSAMVAGVGVVNNINTKLGYNNVAGAGTGIVIDPNGVLT 594
QY      68 NHVVIAGATDIINAFSGSGQTGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
Db      595 NHVVIAGATDIINAFSGSGQTGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 654
QY      128 VVAMGSGGGGTTPRAVGRVVALGTVQASDLSLTAETLNLQIFDAAIQFGDAGGPV 187
Db      655 VVAMGSGGGGTTPRAVGRVVALGTVQASDLSLTAETLNLQIFDAAIQFGDAGGPV 714
QY      188 VNGLGVGVGNNTAAS 202
Db      715 VNGLGVGVGNNTAAS 729

RESULT 9
US-09-287-849-28
/ Sequence 28, Application US/09287849
/ Patent No. 6627198
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Alderson, Mark
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
/ FILE REFERENCE: 014058-009020US
/ CURRENT APPLICATION NUMBER: US/09/287,849
/ CURRENT FILING DATE: 1999-04-07
/ PRIOR APPLICATION NUMBER: US 08/818,112
/ PRIOR FILING DATE: 1997-03-13
/ PRIOR APPLICATION NUMBER: US 08/942,578
/ PRIOR FILING DATE: 1997-10-01
/ PRIOR APPLICATION NUMBER: US 09/025,197
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 09/056,556
/ PRIOR FILING DATE: 1998-04-07
/ PRIOR APPLICATION NUMBER: US 09/223,040
/ PRIOR FILING DATE: 1998-12-30
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 28
/ LENGTH: 231

/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-28

Query Match          39.8%; Score 676; DB 4; Length 231;
Best Local Similarity 99.2%; Pred. No. 3e-51;
Matches 132; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      198 NTAASDNFQLSOGGGAIPIGQAWAIAQIRSGGSPTVHIGPTAFILGLGVVDNNGGA 257
Db      8 HTAASDNFQLSOGGGAIPIGQAWAIAQIRSGGSPTVHIGPTAFILGLGVVDNNGGA 67
QY      258 RVQRVWGSAPAAASLGISTGVDVITAVDGAPINSATAMADALNHHHPGDVISVTWTKSGGT 317
Db      68 RVQRVWGSAPAAASLGISTGVDVITAVDGAPINSATAMADALNHHHPGDVISVTWTKSGGT 127
QY      318 RTGNVTTLAEGPPA 330
Db      128 RTGNVTTLAEGPPA 140

RESULT 10
US-08-818-112-66
/ Sequence 66, Application US/08818112
/ Patent No. 6290969
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
/ NUMBER OF SEQUENCES: 153
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED AND BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/818,112
/ FILING DATE: 13-MAR-1997
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.41106
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 66:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 132 amino acids
/ TYPE: amiro acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-08-818-112-66

Query Match          39.5%; Score 670; DB 3; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 199 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 258
Db 1 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 60
QY 259 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 11
US-08-818-111-67
; Sequence 67, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 210121.417C6
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-67

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 199 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 258
Db 1 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 60
QY 259 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 12
US-09-056-556-66
; Sequence 66, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 210121.457
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-66
Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 199 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 258
Db 1 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 60
QY 259 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 13
US-09-072-596-67
; Sequence 67, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.


```
/ APPLICANT: Hendrickson, Ronald C.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
/ NUMBER OF SEQUENCES: 350
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED AND BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/072,596
/ FILING DATE: 05-MAY-1998
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.417C9
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 67:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 132 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-09-072-596-67

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 199 TAASDNFOLSGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
Db 1 TAASDNFOLSGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
QY 259 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 14
US-09-072-967-66
/ Sequence 66, Application US/09072967
/ Patent No. 6592877
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Hendrickson, Ronald C.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
/ NUMBER OF SEQUENCES: 355
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED AND BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
```

```
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/072,967
/ FILING DATE: 05-MAY-1998
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.411C9
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 66:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 132 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-09-072-967-66

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 199 TAASDNFOLSGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
Db 1 TAASDNFOLSGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
QY 259 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 15
US-09-636-215-819
/ Sequence 819, Application US/09636215
/ Patent No. 6620922
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yuqui
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darick
/ APPLICANT: Li, Samuel
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Hepler, William
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
/ FILE REFERENCE: 210121.42717C17
/ CURRENT APPLICATION NUMBER: US/09/636,215
/ CURRENT FILING DATE: 2000-08-10
/ NUMBER OF SEQ ID NOS: 852
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 819
```

```
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-636-215-819

Query Match      39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 199 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGGFTVHIGPTAFGLGLGVVDNNGNGAR 258
      |||
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGGFTVHIGPTAFGLGLGVVDNNGNGAR 60

Qy 259 VQVVVGSAPAASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISVWQTKSGGTR 318
      |||
Db 61 VQVVVGSAPAASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISVWQTKSGGTR 120

Qy 319 TGNVTLAEGPPA 330
      |||
Db 121 TGNVTLAEGPPA 132

Search completed: June 22, 2004, 17:27:29
Job time : 12.78 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 32,2882 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-12
Perfect score: 1306
Sequence: 1 VAWSVTAGQAEELTAQVRV.....YGHRRGGKYAXSGRRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1304	99.8	263	AAW32379	Mycobacte
2	1304	99.8	263	AAW32447	Mycobacte
3	1304	99.8	263	AAW64317	Mycobacte
4	1304	99.8	263	AAW81680	M. tuberc
5	1304	99.8	263	AAV32062	Mycobacte
6	1304	99.8	263	AAV38984	M. tuberc
7	1304	99.8	263	AAV39121	M. tuberc
8	1304	99.8	263	AAE29706	Mycobacte
9	1304	99.8	263	AAE17570	Mycobacte
10	1304	99.8	358	AAU74591	Antigenic
11	1187	90.9	391	AAW32381	Mycobacte
12	1187	90.9	391	AAW32449	Mycobacte
13	1187	90.9	391	AAW64335	Mycobacte
14	1187	90.9	391	AAW81702	M. tuberc
15	1187	90.9	391	AAV04778	Mycobacte
16	1187	90.9	391	AAV38989	M. tuberc
17	1187	90.9	391	AAV39132	M. tuberc
18	1187	90.9	391	AAU01888	M. tuberc
19	1187	90.9	391	AAE29707	Mycobacte
20	1187	90.9	391	AAE17571	Mycobacte
21	1187	90.9	596	AAV32070	Mycobacte
22	1187	90.9	596	AAE29710	Mycobacte
23	1187	90.9	596	AAE17574	Mycobacte
24	1187	90.9	596	AAU74599	Antigenic
25	1187	90.9	600	AAV32068	Mycobacte

26	1187	90.9	600	5	AAU74597	Antigenic
27	1187	90.9	723	7	ADA26354	Mycobacte
28	1187	90.9	729	4	AAO22142	Ral2-H9-3
29	1187	90.9	729	5	AAE29709	Mycobacte
30	1187	90.9	729	5	AAE17573	Mycobacte
31	1187	90.9	729	7	ADA26374	Mycobacte
32	1187	90.9	744	4	AAU01902	M. tuberc
33	1187	90.9	788	4	AAU01903	M. tuberc
34	1187	90.9	813	7	ADA26367	Mycobacte
35	1187	90.9	815	4	AAU01904	M. tuberc
36	1187	90.9	825	7	ADA26366	Mycobacte
37	1187	90.9	875	7	ADA26365	Mycobacte
38	1187	90.9	930	5	AAE29731	Mycobacte
39	1187	90.9	930	7	ADA26364	Mycobacte
40	1187	90.9	1010	7	ADA26356	Mycobacte
41	1187	90.9	1016	7	ADA26370	M. bovis
42	1187	90.9	1022	7	ADA26369	Mycobacte
43	1187	90.9	1184	7	ADA26368	Mycobacte
44	1182	90.5	394	2	AAU04779	Mycobacte
45	1182	90.5	729	5	AAE29708	Mycobacte

ALIGNMENTS

RESULT 1
AAW32379
ID AAW32379 standard; protein; 263 AA.
XX
AC AAW32379;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-9.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 254
FT /note= "Any amino acid"
XX
FN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
XX
PR 22-SEP-1995; 95US-00532136.
XX
PR 22-MAR-1996; 96US-00620280.
XX
PR 05-JUN-1996; 96US-00658800.
XX
PR 12-JUL-1996; 96US-00680573.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX
DR WPI; 1997-192904/17.
XX
DR N-PSDB; AAT91432.
XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
PT useful for diagnosis of M. tuberculosis infection.
XX
PS Example 3; Page 138-139; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,

CC TbH-9. The immunogenic polypeptide can be used to diagnose M.tuberculosis
 CC infection by forming complexes with specific antibodies in the sample.
 CC Fragments of DNA encoding the immunogenic polypeptide can be used as
 CC diagnostic primers or probes and agents that bind to the antigen,
 CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

XX Sequence 263 AA;
 SQ

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMLIATNLLGQNTA 60
 DB 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMLIATNLLGQNTA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLEQAAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN 180
 DB 121 TAAANQLMNNVPQALKQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN 180
 QY 181 SGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSSLGSSGLGGGVAANLGRAA 240
 DB 181 SGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSSLGSSGLGGGVAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 2

AAW32447
 ID AAW32447 standard; protein; 263 AA.

AC AAW32447;
 DT 09-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen TbH-9.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.

OS Mycobacterium tuberculosis.

Key Location/Qualifiers
 FT Misc-difference 254 /note= "Any amino acid"

PN WO9709428-A2.

PD 13-MAR-1997.

PF 30-AUG-1996; 96WO-US014674.

PR 01-SEP-1995; 95US-00523436.

PR 22-SEP-1995; 95US-00533634.

PR 22-MAR-1996; 96US-00620874.

PR 05-JUN-1996; 96US-00652683.

PR 12-JUL-1996; 96US-00680574.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;

XX Vedvick TH, Twardzik DR;

XX WPI; 1997-192903/17.

XX N-PSDB; AAT91496.

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.

XX Example 3; Page 126-127; 168pp; English.

CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbH-9. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)

XX Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;

Best Local Similarity 100.0%; Pred. No. 3.1e-105;

Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMLIATNLLGQNTA 60

DB 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMLIATNLLGQNTA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLEQAAAVEEASD 120

DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLEQAAAVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN 180

DB 121 TAAANQLMNNVPQALKQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN 180

QY 181 SGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSSLGSSGLGGGVAANLGRAA 240

DB 181 SGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSSLGSSGLGGGVAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263

DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 3

AAW64317
 ID AAW64317 standard; protein; 263 AA.

AC AAW64317;

DT 17-OCT-2003 (revised)

DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen TbH-9.

XX Tuberculosis; infection; diagnosis; antigen; TbH-9.

OS Mycobacterium tuberculosis; strain H37Rv.

Key Location/Qualifiers

FT Misc-difference 254 /note= "encoded by NAG"

PN WO9816645-A2.

PD 23-APR-1998.

PF 07-OCT-1997; 97WO-US018214.

PR 11-OCT-1996; 96US-00729622.

PR 13-MAR-1997; 97US-00618111.

XX (CORI-) CORIXA CORP.

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 DR WPI: 1998-251292/22.
 DR N-PSDB; AAV44371.
 XX
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 XX Example 3; Page 125-126; 250pp; English.
 PS
 XX This polypeptide comprises Mycobacterium tuberculosis antigen TbH-9. A
 CC DNA sequence (see AAV44371) coding for antigen TbH-9 was isolated from a
 CC M. tuberculosis strain H37Rv expression library using sera from patients
 CC having pulmonary or pleural tuberculosis. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAV64291-W64379) comprising an antigenic portion of a
 CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic kits for
 CC detecting M. tuberculosis infection in a patient using these
 CC polypeptides, antibodies or oligonucleotide probes and primers, for the
 CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX
 XX Sequence 263 AA;
 SQ

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
 DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
 QY 181 SGVSWTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGLSGGLGGVAAANLGRAA 240
 DB 181 SGVSWTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGLSGGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4
 AAW81680
 ID AAW81680 standard; protein; 263 AA.
 XX
 AC AAW81680;
 XX
 DT 27-JAN-1999 (first entry)
 XX
 XX M. tuberculosis immunogenic polypeptide TbH-9.
 XX
 DE Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 XX vaccine; pharmaceutical; infection; diagnosis.
 KW
 KW Mycobacterium tuberculosis.
 OS
 XX Key Location/Qualifiers
 XX Misc-difference 254
 FT /label= unknown
 FT
 XX

BN WO9816646-A2.
 XX
 XX 23-APR-1998.
 PD
 XX 07-OCT-1997; 97WO-US018293.
 PF
 XX 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 XX Vedvick TS, Twardzik DR, Lodes MJ;
 PI
 XX WPI: 1998-261042/23.
 DR N-PSDB; AAV64479.
 DR
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 XX Example 3b; Page 119-120; 230pp; English.
 PS
 XX This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX
 XX Sequence 263 AA;
 SQ

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATALLPFEAPEMTSAGGLEQAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATALLPFEAPEMTSAGGLEQAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
 DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
 QY 181 SGVSWTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGLSGGLGGVAAANLGRAA 240
 DB 181 SGVSWTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGLSGGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5
 AAY32062
 ID AAY32062 standard; protein; 263 AA.
 XX
 AC AAY32062;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 XX Mycobacterium tuberculosis antigen TbH9.
 DE
 XX Tuberculosis; antigen; fusion protein; TbH9; diagnosis; therapy; vaccine;
 KW immunogen.
 KW Mycobacterium tuberculosis.
 OS
 XX Key Location/Qualifiers
 XX

FT Misc-difference 254 /note= "not identified"

XX PN WO9951748-A2.

XX PD 14-OCT-1999.

XX PF 07-APR-1999; 99WO-US007717.

XX PR 07-APR-1998; 98US-00056556.

XX PR 30-DEC-1998; 98US-00223040.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky YAW, Alderson M, Campos-Neto A;

XX PI WPI; 1999-601610/51.

XX DR New fusion proteins useful for diagnosis, prevention and treatment of tuberculosis.

XX PT Claim 1; Fig 4C-D; 83pp; English.

XX PS This sequence represents the Mycobacterium tuberculosis antigen TbH9. The

XX CC invention provides fusion proteins (see AAY32059-71) containing at least

XX CC 2 M. tuberculosis antigens such as TbH9, e.g. Mtb32A (see AAY32059) and a

XX CC TBH9-Tb38-1 fusion. The new fusion proteins are useful as vaccines for

XX CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or

XX CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),

XX CC monitoring of disease progression, and treatment of tuberculosis. They

XX CC are more effective immunogens than mixtures of the individual protein

XX CC components

XX SQ Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;

Best Local Similarity 100.0%; Pred. No. 3.1e-105;

Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEILTAACQVRVAAAAYETAYGLTVPPVPIAENRAELMILITATNLLGQNTPA 60

Db 1 VAMSVTAGQAEILTAACQVRVAAAAYETAYGLTVPPVPIAENRAELMILITATNLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGLLLEQAAVEEASD 120

Db 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGLLLEQAAVEEASD 120

QY 121 TAAANQLMNNVPOALKOLAQPTGTTTSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180

Db 121 TAAANQLMNNVPOALKOLAQPTGTTTSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180

QY 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSGLGGVAAANLGRAA 240

Db 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263

Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 6

AAY38984

ID AAY38984 standard; protein; 263 AA.

XX AC AAY38984;

XX DT 05-NOV-1999 (first entry)

XX M. tuberculosis recombinant antigen protein TbH-9.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;

XX vaccine; immunity.

XX

OS Mycobacterium tuberculosis.

XX PN WO9942118-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US003265.

XX PR 18-FEB-1998; 98US-00024753.

XX PR 05-MAY-1998; 98US-00072596.

XX PA (CORI-) CORIXA CORP.

XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

XX PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;

XX DR WPI; 1999-527416/44.

XX DR N-PSDB; AAZ19069.

XX PT New polypeptide comprising antigenic portions of M. tuberculosis.

XX PS Example 3; Page 160-161; 323pp; English.

XX CC This invention describes novel recombinant antigens and their encoding

XX CC nucleic acids derived from Mycobacterium tuberculosis. The novel

XX CC polypeptides are useful for detecting M. tuberculosis infection in a

XX CC biological sample by detecting antibodies which bind with the

XX CC polypeptides, and are useful as vaccines for immunizing against M.

XX CC tuberculosis infection. The new detection methods are needed as current

XX CC vaccination strategies do not provide 100% immunity

XX SQ Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;

Best Local Similarity 100.0%; Pred. No. 3.1e-105;

Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEILTAACQVRVAAAAYETAYGLTVPPVPIAENRAELMILITATNLLGQNTPA 60

Db 1 VAMSVTAGQAEILTAACQVRVAAAAYETAYGLTVPPVPIAENRAELMILITATNLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPEMTSAGLLLEQAAVEEASD 120

Db 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPEMTSAGLLLEQAAVEEASD 120

QY 121 TAAANQLMNNVPOALKOLAQPTGTTTSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180

Db 121 TAAANQLMNNVPOALKOLAQPTGTTTSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180

QY 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSGLGGVAAANLGRAA 240

Db 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263

Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 7

AAY39121

ID AAY39121 standard; protein; 263 AA.

XX AC AAY39121;

XX DT 05-NOV-1999 (first entry)

XX M. tuberculosis antigen TbH-9 amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;

XX immunotherapy; diagnosis; immunisation; vaccine; infection;

XX immune response; skin test.

XX

OS Mycobacterium tuberculosis.

```

XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US003268.
XX PR 18-FEB-1998; 98US-00025197.
XX PR 05-MAY-1998; 98US-00072967.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX DR WPI; 1999-527409/44.
XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
XX PT tests and protective or therapeutic vaccines or compositions.
XX PS Example 3; Page 115; 299pp; English.
XX CC The present invention describes polypeptides comprising an immunogenic
XX CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
XX CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
XX CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
XX CC polypeptides fragments, can be used in pharmaceutical compositions or
XX CC vaccines to generate a protective or therapeutic immune response to M.
XX CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
XX CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
XX CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
XX CC to AA219460 and AA39083 to AA39225 are used in the exemplification of
XX CC the present invention
XX SQ Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.1e-105; Indels 0; Gaps 0;
Matches 263; Conservative 0; Mismatches 0;

QY 1 VAMSVTAGQAEATAAQRVAAAAYETAYGLTVPVPIAENRAELMILITNLLGQNTPA 60
DB 1 VAMSVTAGQAEATAAQRVAAAAYETAYGLTVPVPIAENRAELMILITNLLGQNTPA 60

QY 61 IAVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 61 IAVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180

QY 181 SGVSMNTLTSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLTSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 8
AAE29706
ID AAE29706 standard; protein; 263 AA.
XX AC AAE29706;
XX DT 27-JAN-2003 (first entry)
XX DE Mycobacterium sp. TbH9 antigenic protein.
XX KW Vaccine; immunity; diagnostic agent; gene therapy; TbH9 antigen.
XX
```

```

OS Mycobacterium sp.
XX Key Location/Qualifiers
XX FT Misc-difference 254
XX FT /note= "Encoded by NAG"
XX PN WO200272792-A2.
XX PD 19-SEP-2002.
XX PF 13-MAR-2002; 2002WO-US008223.
XX PR 13-MAR-2001; 2001US-0275837P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Brannon M, Guderian J;
XX DR WPI; 2002-759844/82.
XX DR N-PSDB; AAD47081.
XX PT New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
XX PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX PT tuberculosis.
XX PS Disclosure; Page 84; 155pp; English.
XX CC The invention relates to a recombinant nucleic acid molecule encoding a
XX CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX CC polynucleotide sequence encoding an antigen or an antigenic fragment from
XX CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX CC polypeptide or its fragment. The Leishmania polynucleotide is selected
XX CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
XX CC are used in methods for eliciting immune response in mammals. They are
XX CC useful as vaccines to elicit protective immunity against pathogenic
XX CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX CC polypeptides are used for enhancing the expression of polynucleotides, as
XX CC in vivo diagnostic agents and for raising antibodies in a non-human is
XX CC animal. The invention is used in gene therapy. The present sequence is
XX CC Mycobacterium sp. TBH9 antigenic protein
XX SQ Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 5; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.1e-105; Indels 0; Gaps 0;
Matches 263; Conservative 0; Mismatches 0;

QY 1 VAMSVTAGQAEATAAQRVAAAAYETAYGLTVPVPIAENRAELMILITNLLGQNTPA 60
DB 1 VAMSVTAGQAEATAAQRVAAAAYETAYGLTVPVPIAENRAELMILITNLLGQNTPA 60

QY 61 IAVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 61 IAVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180

QY 181 SGVSMNTLTSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLTSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 9
AAE17570
ID AAE17570 standard; protein; 263 AA.
XX
```

AC AAE17570;
 XX 22-APR-2002 (first entry)
 DT
 XX Mycobacterium species MTB39 (TbH9) protein #1.
 DE
 XX Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB39; TbH9 protein.
 XX
 XX Mycobacterium sp.
 OS
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 254
 FT /label= Unknown
 FT /note= "Encoded by NAG"
 FT
 XX WO200198460-A2.
 XX
 XX 27-DEC-2001.
 PD
 XX 20-JUN-2001; 2001WO-US019959.
 PF
 XX 20-JUN-2000; 2000US-00597796.
 PR
 XX 01-FEB-2001; 2001US-0265737P.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX
 XX Skeiky Y, Reed S, Alderson M;
 PI
 XX WPI; 2002-147798/19.
 DR
 XX N-PSDB; AAD28340.
 DR
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 PT
 XX Claim 83; Page 100; 136pp; English.
 PS
 XX
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB39 (TbH9) protein
 XX
 XX Sequence 263 AA;
 SQ
 Query March 99.8%; Score 1304; DB 5; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3,1e-105; Gaps 0;
 Matches 263; Conservative 0; Mismatches 0; Indels 0;
 QY 1 VAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIATNLLGNTPA 60
 Db 1 VAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIATNLLGNTPA 60
 QY 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPMTSAGLLLEQAAAYEASD 120
 Db 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPMTSAGLLLEQAAAYEASD 120

QY 121 TAAANQIMNNVPQALKQLAQAOTGTTPTSSKLGGLWKTVPSPHSPISNMVSNNHMSMTN 180
 Db 121 TAAANQIMNNVPQALKQLAQAOTGTTPTSSKLGGLWKTVPSPHSPISNMVSNNHMSMTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGSSGSLGGGVAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGSSGSLGGGVAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 RESULT 10
 AAU74591
 ID AAU74591 standard; protein; 358 AA.
 XX
 AC AAU74591;
 XX
 DT 29-AUG-2003 (revised)
 DT 08-MAY-2002 (first entry)
 XX
 DE Antigenic fusion protein TbH9-Tb38-1.
 XX
 KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; TbH9-Tb38-1; TbH9; Tb38-1.
 XX
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 254
 FT /label= unknown
 XX
 PN US2002009459-A1.
 XX
 PD 24-JAN-2002.
 XX
 XX 07-APR-1999; 99US-00287849.
 PR 13-MAR-1997; 97US-00818112.
 PR 01-OCT-1997; 97US-00942578.
 PR 18-FEB-1998; 98US-00025197.
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 PA (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.
 XX
 PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 XX WPI; 2002-171134/22.
 XX
 XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis.
 XX
 PS Claim 1; Fig 4C-D; 62pp; English.
 XX
 CC The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention. Note: The specification states that this polypeptide is
 CC encoded by the polynucleotide shown in ABK14131. (Updated on 29-AUG-2003


```

XX WPI: 1998-261042/23.
DR N-PSDB; AAV64503.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
PT develop products for the detection of M. tuberculosis infection and for
PT diagnosis, treatment and prevention of tuberculosis.
XX
XX Example 3B; Page 128-129; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
CC inducing protective immunity against tuberculosis (TB). This sequence can
CC be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis
XX
XX Sequence 391 AA;
  Query Match      90.9%; Score 1187; DB 2; Length 391;
  Best Local Similarity 99.6%; Pred. No. 8.2e-95;
  Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMILITNLLGQNTPA 60
DB 74 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMILITNLLGQNTPA 133
QY 61 IAVNEAEGEWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 134 IAVNEAEGEWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 193
QY 121 TAAANQLMNNVPQALKQLAQPTGTTSPSSKLGGLWKTVPSPHRSPISNMYSMANNHSMTN 180
DB 194 TAAANQLMNNVPQALKQLAQPTGTTSPSSKLGGLWKTVPSPHRSPISNMYSMANNHSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA 240
DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA 313
QY 241 SV 242
DB 314 SV 315

```

```

RESULT 15
AAY04778
ID AAY04778 standard; protein; 391 AA.
XX
AC AAY04778;
XX
XX 06-JUL-1999 (first entry)
XX
DE Mycobacterium species protein sequence 5R.
XX
XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection.
XX
OS Mycobacterium sp.
XX
XX WO9909186-A2.
XX
PD 25-FEB-1999.
XX
PF 14-AUG-1998; 98WO-FR001813.
XX
PR 14-AUG-1997; 97FR-00010404.
PR 11-SEP-1997; 97FR-00011325.
XX
XX (INSP ) INST PASTEUR.
XX
XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
PI Goguet De La Salmoniere Y;
XX

```

```

DR WPI: 1999-181045/15.
DR N-PSDB; AAX34030.
XX

```

```

PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
PT protein expression.
XX

```

```

PS Claim 32; Fig 5R; 309pp; French.
XX

```

```

CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
CC from various Mycobacterium species microorganisms. The encoding
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection
XX

```

```

XX Sequence 391 AA;
  Query Match      90.9%; Score 1187; DB 2; Length 391;
  Best Local Similarity 99.6%; Pred. No. 8.2e-95;
  Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMILITNLLGQNTPA 60
DB 74 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMILITNLLGQNTPA 133
QY 61 IAVNEAEGEWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 134 IAVNEAEGEWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 193
QY 121 TAAANQLMNNVPQALKQLAQPTGTTSPSSKLGGLWKTVPSPHRSPISNMYSMANNHSMTN 180
DB 194 TAAANQLMNNVPQALKQLAQPTGTTSPSSKLGGLWKTVPSPHRSPISNMYSMANNHSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA 240
DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA 313
QY 241 SV 242
DB 314 SV 315

```

```

Search completed: June 22, 2004, 17:15:28
Job time : 33.2882 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 9.38833 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-12

Perfect score: 1306

Sequence: 1 VAMSVTAQAEILTAQVRV.....YGHDDGGKYAXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptcdat2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptcdat2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptcdat2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptcdat2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptcdat2/iaa/pttus_COMB.pep.*
6: /cgn2_6/ptcdat2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1304	99.8	263	3	US-08-818-112-91
2	1304	99.8	263	4	US-08-818-111-92
3	1304	99.8	263	4	US-09-056-556-91
4	1304	99.8	263	4	US-09-072-596-92
5	1304	99.8	263	4	US-09-072-967-91
6	1304	99.8	358	4	US-09-287-849-8
7	1187	90.9	391	3	US-08-818-112-107
8	1187	90.9	391	4	US-08-818-111-102
9	1187	90.9	391	4	US-09-056-556-107
10	1187	90.9	391	4	US-09-072-596-102
11	1187	90.9	391	4	US-09-072-967-107
12	1187	90.9	596	4	US-09-287-849-26
13	1187	90.9	600	4	US-09-287-849-22
14	1182	90.5	729	4	US-09-223-040-2
15	1182	90.5	729	4	US-09-287-849-2
16	1001.5	76.7	396	3	US-08-818-112-111
17	1001.5	76.7	396	4	US-08-818-111-106
18	1001.5	76.7	396	4	US-09-056-556-111
19	1001.5	76.7	396	4	US-09-072-596-106
20	1001.5	76.7	396	4	US-09-072-967-111
21	949.5	72.7	359	3	US-08-818-112-109
22	949.5	72.7	359	4	US-08-818-111-104
23	949.5	72.7	359	4	US-09-056-556-109
24	949.5	72.7	359	4	US-09-072-596-104
25	949.5	72.7	359	4	US-09-072-967-109
26	429.5	32.9	400	4	US-09-073-009-126
27	429.5	32.9	400	4	US-09-073-010-126

28 325.5 24.9 423 4 US-09-073-009-142 Sequence 142, App
29 325.5 24.9 423 4 US-09-073-010-142 Sequence 142, App
30 325.5 24.9 710 4 US-09-287-849-16 Sequence 16, Appl
31 325.5 24.9 856 4 US-09-287-849-12 Sequence 12, Appl
32 320.5 24.5 943 4 US-09-477-135A-131 Sequence 131, Appl
33 239 18.3 204 4 US-08-311-731A-57 Sequence 57, Appl
34 235 18.0 208 4 US-08-311-731A-208 Sequence 208, App
35 223 17.1 141 4 US-09-073-009-15 Sequence 15, Appl
36 223 17.1 141 4 US-09-073-010-15 Sequence 15, Appl
37 197.5 15.1 943 4 US-09-056-556-204 Sequence 204, App
38 197.5 15.1 943 4 US-09-072-596-199 Sequence 199, App
39 197.5 15.1 943 4 US-09-072-967-204 Sequence 204, App
40 188.5 14.4 368 3 US-08-818-112-114 Sequence 114, App
41 188.5 14.4 368 4 US-08-818-111-109 Sequence 109, App
42 188.5 14.4 368 4 US-09-056-556-114 Sequence 114, App
43 188.5 14.4 368 4 US-09-072-596-109 Sequence 109, App
44 188.5 14.4 368 4 US-09-072-967-114 Sequence 114, App
45 188.5 14.4 371 4 US-09-050-739-92 Sequence 92, Appl

ALIGNMENTS

RESULT 1

US-08-818-112-91
; Sequence 91, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-91

Query Match 99.8%; Score 1304; DB 3; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 TAAANQLMNNVPQALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Qy 181 SGVSMNTLTSSMLKGFAPAAAQAVQTAACNGVRAMSSSLGSSGLGGVAAANLGRAA 240
Db 181 SGVSMNTLTSSMLKGFAPAAAQAVQTAACNGVRAMSSSLGSSGLGGVAAANLGRAA 240
Qy 241 SVRYGHRDGGKYAXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4
US-09-072-596-92
; Sequence 92, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-92

Query Match 99.8%; Score 1304; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VAWSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMILLIATNLLGNTPA 60
Db 1 VAWSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMILLIATNLLGNTPA 60
Qy 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLEQAAAVEASD 120
Db 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLEQAAAVEASD 120
Qy 121 TAAANQLMNNVPQALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Db 121 TAAANQLMNNVPQALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180

Qy 181 SGVSMNTLTSSMLKGFAPAAAQAVQTAACNGVRAMSSSLGSSGLGGVAAANLGRAA 240
Db 181 SGVSMNTLTSSMLKGFAPAAAQAVQTAACNGVRAMSSSLGSSGLGGVAAANLGRAA 240
Qy 241 SVRYGHRDGGKYAXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5
US-09-072-967-91
; Sequence 91, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-91

Query Match 99.8%; Score 1304; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VAWSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMILLIATNLLGNTPA 60
Db 1 VAWSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMILLIATNLLGNTPA 60
Qy 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLEQAAAVEASD 120
Db 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLEQAAAVEASD 120
Qy 121 TAAANQLMNNVPQALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Db 121 TAAANQLMNNVPQALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180

Qy	121	TAANQLMNVPQALKOLAQPTQGTTPSSKLGGLWKTVP	PHRSPISNMVSWANNHSMTN	180
Db	335	TAANQLMNVPQALQQAQPTQGTTPSSKLGGLWKTVP	PHRSPISNMVSWANNHSMTN	394
Qy	181	SGVSVNTLTLSMLKGFAFAAAQAQVOTAAQNGVFRAM	SLGSSLGSSGLGGGVAANLGRAA	240
Db	395	SGVSVNTLTLSMLKGFAFAAAQAQVOTAAQNGVFRAM	SLGSSLGSSGLGGGVAANLGRAA	454
Qy	241	SV	242	
Db	455	SV	456	

```

RESULT 15
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:trf-fusion
US-09-287-849-2

```

Search completed: June 22, 2004, 17:27:31
Job time : 9.38833 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 24.3589 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-12
Perfect score: 1306
Sequence: 1 VAWMSVTAGQBELTAQVRV.....YCHRDGKYAXSGRRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
-Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1304	99.8	263	12	US-09-886-349A-12
2	1304	99.8	263	14	US-10-193-002-92
3	1304	99.8	263	14	US-10-084-843-91
4	1304	99.8	263	14	US-10-098-732A-12
5	1304	99.8	358	9	US-09-287-849-8
6	1304	99.8	358	14	US-10-359-460-8
7	1187	90.9	391	12	US-09-886-349A-14
8	1187	90.9	391	14	US-10-193-002-102
9	1187	90.9	391	14	US-10-084-843-107
10	1187	90.9	391	14	US-10-098-732A-14
11	1187	90.9	596	9	US-09-287-849-26
12	1187	90.9	596	12	US-09-886-349A-20
13	1187	90.9	596	14	US-10-359-460-26
14	1187	90.9	596	14	US-10-098-732A-20
15	1187	90.9	600	9	US-09-287-849-22

16	1187	90.9	600	14	US-10-359-460-22	Sequence 22, Appl
17	1187	90.9	723	15	US-10-369-983-2	Sequence 2, Appl
18	1187	90.9	723	12	US-09-886-349A-18	Sequence 18, Appl
19	1187	90.9	729	14	US-10-098-732A-18	Sequence 18, Appl
20	1187	90.9	729	15	US-10-369-983-21	Sequence 21, Appl
21	1187	90.9	729	15	US-10-369-983-22	Sequence 22, Appl
22	1187	90.9	813	15	US-10-369-983-15	Sequence 15, Appl
23	1187	90.9	825	15	US-10-369-983-14	Sequence 14, Appl
24	1187	90.9	875	15	US-10-369-983-13	Sequence 13, Appl
25	1187	90.9	930	14	US-10-098-732A-65	Sequence 65, Appl
26	1187	90.9	930	15	US-10-369-983-12	Sequence 12, Appl
27	1187	90.9	1010	15	US-10-369-983-4	Sequence 4, Appl
28	1187	90.9	1016	15	US-10-369-983-18	Sequence 18, Appl
29	1187	90.9	1022	15	US-10-369-983-17	Sequence 17, Appl
30	1187	90.9	1154	15	US-10-369-983-16	Sequence 16, Appl
31	1182	90.5	729	9	US-09-287-849-2	Sequence 2, Appl
32	1182	90.5	729	12	US-09-886-349A-16	Sequence 16, Appl
33	1182	90.5	729	14	US-10-359-460-2	Sequence 2, Appl
34	1182	90.5	729	14	US-10-098-732A-16	Sequence 16, Appl
35	1182	90.5	729	15	US-10-359-459-2	Sequence 2, Appl
36	1179	90.3	391	12	US-09-872-186-8	Sequence 106, App
37	1001.5	76.7	396	14	US-10-193-002-106	Sequence 111, App
38	1001.5	76.7	396	14	US-10-084-843-111	Sequence 104, App
39	949.5	72.7	359	14	US-10-193-002-104	Sequence 109, App
40	949.5	72.7	359	14	US-10-084-843-109	Sequence 62455, A
41	949.5	72.7	393	12	US-10-282-122A-62455	Sequence 64892, A
42	949.5	72.7	393	12	US-10-282-122A-64892	Sequence 126, App
43	429.5	32.9	400	9	US-09-073-009-126	Sequence 126, App
44	429.5	32.9	400	9	US-09-793-306-126	Sequence 62027, A
45	429.5	32.9	405	12	US-10-282-122A-62027	

ALIGNMENTS

RESULT 1
US-09-886-349A-12
; Sequence 12, Application US/09886349A
; Publication No. US2004008623A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTP39 (TbH9)
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid

Query Match 99.8%; Score 1304; DB 12; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAWMSVTAGQBELTAQVRVAAAYAYTAYGLVTPPPVIAENRAELMILIAINLLGQNTPA 60
DB 1 VAWMSVTAGQBELTAQVRVAAAYAYTAYGLVTPPPVIAENRAELMILIAINLLGQNTPA 60

Query Match 99.8%; Score 1304; DB 14; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120

QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 240

QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYXSGRRNGGPA 263

RESULT 4
US-10-098-732A-12
; Sequence 12, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 12
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 (TbH9)
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-10-098-732A-12

Query Match 99.8%; Score 1304; DB 14; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120

QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 240

DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 240
QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYXSGRRNGGPA 263

RESULT 5
US-09-287-849-8
; Sequence 8, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Devin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
; OTHER INFORMATION: protein TbH9-Tb38-1
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-287-849-8

Query Match 99.8%; Score 1304; DB 9; Length 358;
Best Local Similarity 100.0%; Pred. No. 1.7e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120

QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 240

QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYXSGRRNGGPA 263

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-10-193-002-102

Query Match 90.9%; Score 1187; DB 14; Length 391;
Best Local Similarity 99.6%; Pred. No. 3.8e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAQGAELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAITNLLQNTPA	60
Db	74	VAMSVTAQGAELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAITNLLQNTPA	133
QY	61	IAVNEAEGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGGLLEQAAVEEASD	120
Db	134	IAVNEAEGEMWAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAVEEASD	193
QY	121	TAAANQNMNVPAQLKQALQAOPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMTN	180
Db	194	TAAANQNMNVPAQLKQALQAOPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMTN	253
QY	181	SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSGLGGGVAANLGRAA	240
Db	254	SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSGLGGGVAANLGRAA	313
QY	241	SV 242	
Db	314	SV 315	

RESULT 9
US-10-084-843-107
Sequence 107, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 107:
US-10-084-843-107

Query Match 90.9%; Score 1187; DB 14; Length 391;
Best Local Similarity 99.6%; Pred. No. 3.8e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAQGAELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAITNLLQNTPA	60
Db	74	VAMSVTAQGAELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAITNLLQNTPA	133
QY	61	IAVNEAEGEMWAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAVEEASD	120
Db	134	IAVNEAEGEMWAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAVEEASD	193
QY	121	TAAANQNMNVPAQLKQALQAOPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMTN	180
Db	194	TAAANQNMNVPAQLKQALQAOPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMTN	253
QY	181	SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSGLGGGVAANLGRAA	240
Db	254	SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSGLGGGVAANLGRAA	313
QY	241	SV 242	
Db	314	SV 315	

RESULT 10
US-10-098-732A-14
Sequence 14, Application US/10098732A
Publication No. US20030175294A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs comprising a
TITLE OF INVENTION: Leishmania Antigen
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 60/275,837
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 full length (TBH9FL)
US-10-098-732A-14

Query Match 90.9%; Score 1187; DB 14; Length 391;
Best Local Similarity 99.6%; Pred. No. 3.8e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAATNLLGQNTPA 60
DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAATNLLGQNTPA 133

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 193

QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 194 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
DB 254 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 313

QY 241 SV 242
DB 314 SV 315

RESULT 11
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 90.9%; Score 1187; DB 9; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAATNLLGQNTPA 60
DB 82 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAATNLLGQNTPA 141

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
DB 142 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 201

QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 202 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 261

QY 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
DB 262 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 321

QY 241 SV 242
DB 322 SV 323

RESULT 12
US-09-886-349A-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59F)
US-09-886-349A-20

Query Match 90.9%; Score 1187; DB 12; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAATNLLGQNTPA 60
DB 82 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAATNLLGQNTPA 141

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
DB 142 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 201

QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 202 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 261

QY 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
DB 262 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 321

QY 241 SV 242
DB 322 SV 323

RESULT 13
US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US2003014791A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-26

Query Match 90.9%; Score 1187; DB 14; Length 596;
Best Local Similarity 99.6%; Pred.No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEHTAAQVRVAAAAAYETAYGLTVPVPIAENRAELMILITNLGQNTPA 60
DB 82 VAMSVTAGQAEHTAAQVRVAAAAAYETAYGLTVPVPIAENRAELMILITNLGQNTPA 141
QY 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATATLLPPEEAPMTSAGLLEQAAAVEEASD 120
DB 142 IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGLLEQAAAVEEASD 201
QY 121 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
DB 202 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 261
QY 181 SGVSMNTTLLSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
DB 262 SGVSMNTTLLSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSGLGGVAAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 14
US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US2003017529A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59F)
US-10-098-732A-20

Query Match 90.9%; Score 1187; DB 14; Length 596;
Best Local Similarity 99.6%; Pred.No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEHTAAQVRVAAAAAYETAYGLTVPVPIAENRAELMILITNLGQNTPA 60
DB 82 VAMSVTAGQAEHTAAQVRVAAAAAYETAYGLTVPVPIAENRAELMILITNLGQNTPA 141
QY 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGLLEQAAAVEEASD 120
DB 142 IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGLLEQAAAVEEASD 201
QY 121 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
DB 202 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 261
QY 181 SGVSMNTTLLSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
DB 262 SGVSMNTTLLSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSGLGGVAAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 15
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-22

Query Match 90.9%; Score 1187; DB 9; Length 600;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQAE	LTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIA	TNLLGQNTPA	60
Db	82	VAMSVTAGQAE	LTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIA	TNLLGQNTPA	141
QY	61	IADVNEA	YGENWAGDAAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAA	VEEASD	120
Db	142	IADVNEA	YGENWAGDAAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAA	VEEASD	201
QY	121	TAAANQLM	NNVPOALKOLAQPTQGTTPSSKLGGLWKTVP	SHRSPISNNVSMANNHMSMTN	180
Db	202	TAAANQLM	NNVPOALKOLAQPTQGTTPSSKLGGLWKTVP	SHRSPISNNVSMANNHMSMTN	261
QY	181	SGVSM	TNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSS	LGSSGLGGVAA	240
Db	262	SGVSM	TNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSS	LGSSGLGGVAA	321
QY	241	SV	242		
Db	322	SV	323		

Search completed: June 22, 2004, 18:07:54
Job time : 24.3589 secs

C:Accession: H70741
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: H70741
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-396 <COL>
 A: Cross-references: GB:Z75555; GB:AL123456; NID:G3261608; PIDN:CAA99966.1; PID:e250360;
 A: Experimental source: strain H37Rv
 C: Genetics:
 A: Gene: PPE

Query Match 76.7%; Score 1001.5; DB 2; Length 396;
 Best Local Similarity 84.1%; Pred. No. 3.9e-61;
 Matches 207; Conservative 13; Mismatches 21; Indels 5; Gaps 2;
 QY 1 VAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 60
 DB 74 VAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEEAPMTSAGGLLEQAAVEEASD 120
 DB 134 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEEAPMTSAGGLLEQAAVEEASD 193
 QY 121 TAAANOLMNNVPOALKOLAQTOGTTPSSKLGGLWKTVSPHRSPISNMVMANNHSMWN 180
 DB 134 TAAANOLMNNVPOALKOLAQTOGTTPSSKLGGLWKTVSPHRSPISNMVMANNHSMWN 253
 QY 181 SGVSMNTLSSMLKGFAPAAQAQVOTAAQNGVRAMS---LGSLSGSSGLGGVAAANL 236
 DB 254 SGVSMNTLSSMLKGFAPAAQAQVOTAAQNGVRAMS---LGSLSGSSGLGGVAAANL 312
 QY 237 GRAASV 242
 DB 313 GRAASV 318

RESULT 3
 C70568
 Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: C70568
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: C70568
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-393 <COL>
 A: Cross-references: GB:Z95390; GB:AL123456; NID:G3261766; PIDN:CAB08702.1; PID:e316074;
 A: Experimental source: strain H37Rv
 C: Genetics:
 A: Gene: PPE

Query Match 72.7%; Score 949.5; DB 2; Length 393;
 Best Local Similarity 79.7%; Pred. No. 1.4e-57;
 Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;
 QY 1 VAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 60
 DB 74 VAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEEAPMTSAGGLLEQAAVEEASD 120

DB 134 IEANQAYSQWQDAEMGYAATAATATEALLPFEDAPLTNPGLLEQAAVEEAD 193
 QY 121 TAAANOLMNNVPOALKOLAQTOGTTPSSKLGGLWKTVSPHRSPISNMVMANNHSMWN 180
 DB 194 TAAANOLMNNVPOALKOLAQPAQGVVPSKLGGLTAVSPHLSPLSNVSSIANNNHSMWG 253
 QY 181 SGVSMNTLSSMLKGFAPAAQAQVOTAAQNGVRAMS---LGSLSGSSGLGGVAAANL 236
 DB 254 TGVSMNTLHSMKGLAP-AAQAQVETAAENGWVWSSLSGSLGSSGLGGVAAANL 312
 QY 237 GRAASV 242
 DB 313 GRAASV 318

RESULT 4
 H70931
 Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: H70931
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: H70931
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-403 <COL>
 A: Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN:CAAL7728.1; PID:e125461f
 A: Experimental source: strain H37Rv
 C: Genetics:
 A: Gene: PPE

Query Match 35.2%; Score 460; DB 2; Length 403;
 Best Local Similarity 43.3%; Pred. No. 3.4e-24;
 Matches 117; Conservative 36; Mismatches 73; Indels 44; Gaps 8;
 QY 1 VAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 60
 DB 77 VAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 136
 QY 61 IAVNEAEYGEWMAQDAAMFGVAAATATATLTPFEEAPMTSAGGLLEQAAVEEASD 120
 DB 137 IAVNEAEYGEWMAQDAAMFGVAAATATATLTPFEEAPMTSAGGLLEQAAVEEASD 195
 QY 121 TAA-----NOLMNNVPOALKOLAQTOGT-----TPSSKLGGLWKTVSPH 161
 DB 196 ASAGAQQTTLSQLIAAIPSVLGGSSSTAATFASGPSGLLGIVGSSGLWKLWALLDPN 255
 QY 162 RSPISNMVMANNHSMWNVSM-TNTLS---SMLKGFAPAAQAQVOTAAQNGVRAMS 217
 DB 256 -----SNFWNTIASSGLFIPSNITAFPLGLGVAADAAGDVLSGTS----- 300
 QY 218 SLGSSL-----GSSGLGGVAAANLGRAASV 242
 DB 301 GLGGALVAPLGSAGGLGCTVAAGLGNAAV 330

RESULT 5
 B70931
 Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: B70931
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

A;Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; PMID:98295987; PMID:9634230
A;Accession: B70931
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-423 <COL>
A;Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17722.1; PID:e125461
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: PPE

Query Match 33.8%; Score 441; DB 2; Length 423;
Best Local Similarity 41.2%; Pred. No. 7e-23;
Matches 117; Conservative 30; Mismatches 67; Indels 70; Gaps 9;
QY 1 VAMSVTAGAQLTAQVRVAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
DB 72 VAMSVATVHAELAGAAQLAIAAYEAAFAATPPPPVIAANRAQLMVLIAITNFGQNTPA 131
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPPEAPEMTSAGGLLEQAAAVEASD 120
DB 132 IMTEAQYMEWMAQDAAMFYAGSSATA-SRTAFTEPQTTHGQLGAQSSAVACTAA 190
QY 121 TAAAN-----QLMNVVPAQLKOLAQPT---QGTFP-----SSKLGGLWKTVP 160
DB 191 TAAGGNLSAFPQLLSAVPRALQGLALPTASQASATPQWVDLGNLSTFLGG-AVTGP 248
QY 161 HRSPISNVSMANNHSMNTSGVMTNTLSMLKGFAPAPAAAQAVTAQAQNGVRAMSSLG 220
DB 249 YTFP-----GVLPPSGVPYLLGIQSVL-----VTQNGQGVSAALG 283
QY 221 S-----SLGSSGLGGG-VAANLGRAASV 242
DB 284 KIGGKPTGALAPLAEFALHTPILGSEGLGGVSAGIGRAGLV 327

RESULT 6
G70929
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: G70929
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; PMID:98295987; PMID:9634230
A;Accession: G70929
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-393 <COL>
A;Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17711.1; PID:e125461
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: PPE

Query Match 33.1%; Score 432.5; DB 2; Length 393;
Best Local Similarity 41.0%; Pred. No. 2.4e-22;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
QY 1 VAMSVTAGAQLTAQVRVAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
DB 73 VAMSVAAAQAQEAQTAQRAAAAFEAFAATPPPLIAANRASLMQLISTNVFGQNTSA 132
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGGLLEQAAAVEASD 120
DB 133 IAAAEAYGEWMAQDSAMVAYAGSSASASA-VTFSTPQIANPRAQGTQAAAVATAG 191
QY 121 TAAAN--NOLMNVVPAQLKOLAQPTQGTTPSSKLGGLWK-----TVSPH 161

DB 192 TAQSTILTEMITGLPNALQSLTSPILQSS-NGPLSLMWQLFGTFNPTTSISALLTDLQPY 250
QY 162 RSPINSMVSMANNHSMNTSGVMTNTLSMLKGFAPAPAAAQAVTAQAQNGVRAMSSLG 221
DB 251 ASFFYNTEGLPYFSIGMGNFTQSARKTL-GLIGSAAPAAVA-----AAGDAAKGLPGLGG 304
QY 222 SLGSSGLGGVAAANLGRAASV 242
DB 305 MLG-----GGPVAAGLGNAAASV 321

RESULT 7
H87056
PPE-family protein [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: H87056
R;Cole, S.T.; Eglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Seeger, K.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Taylor, K.; Whitehead, S.; Barrell, B.G.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; PMID:21128732; PMID:11234002
A;Accession: H87056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-421 <STO>
A;Cross-references: GB:AL450380; NID:g13093150; PIDN:CAC31563.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML1182

Query Match 32.5%; Score 424.5; DB 2; Length 421;
Best Local Similarity 38.8%; Pred. No. 9.3e-22;
Matches 104; Conservative 36; Mismatches 99; Indels 29; Gaps 5;
QY 3 WMSVTAGAQLTAQVRVAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 62
DB 75 WLTQNAASAEALTATQLTVAANAYETAFTMTVPPLMVFVNRAQACILLMSNIFGQNSTA 134
QY 63 VNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGGL----- 108
DB 135 EKEAYETENWIDAAAMTSYQASVLEAVGATKATAPLGVNEVGLAEVVEEVEV 194
QY 109 --LEQAAAVEASD--TAANQLMNN-----VPAQLKQ---LAQPTQGTTPSSKLG--L 154
DB 195 EVVEEVEAEQAISQAALDQAVNEGWEATVVPQVDQVNVDTATPQAVPDSSSAAAPQL 254
QY 155 WKTVPSPHRSPISNVMYMANHSMNTSGVMTNTLSMLKGFAPAPAAAQAVTAQAQNGVR 214
DB 255 WGGFAQHLSPIINDTLISMINNHAGMANAGLSLVNGMSGAMKSLAP-TTTKAAESAPKAMGS 313
QY 215 AMSLSGLSSILGSSGLGGVAAANLGRAASV 242
DB 314 AVQSTGRGLLSSSGGHVTAQLGRAASI 341

RESULT 8
A70932
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: A70932
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; PMID:98295987; PMID:9634230
A;Accession: A70932
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

A:Residues: 1-409 <COL>
A:Cross-references: GB:AL0202021; GB:AL123456; NID:93250699; PIDN:CAAL17729.1; PID:e125461
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 32.3%; Score 422; DB 2; Length 409;
Best Local Similarity 42.2%; Pred. No. 1.3e-21;
Matches 108; Conservative 34; Mismatches 52; Indels 22; Gaps 8;
Db 1 VAMSVTAQAGBELTAQVRAAAAYETAYGLTVPVPIAENRAELMILIAITNLGQNTPA 60
73 VAWISVTAGAEQAGQAQAKIAGVYETAFATVPPVPIAENRALLMSIVATNIFGQNTPA 132
Qy 61 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPPEEAPMTSAGGLLEQAAVEEASD 120
Db 133 IAAEAEQAEWMAQDAAMFGVAAATATATATLTPPEEAPMTSAGGLLEQAAVEEASD 191
Qy 121 TAAA-----NQLMNVNPOALKQAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSAN 173
Db 192 AAASDDITAQLSQLSLPSTLQSLA--TTATATSAG--WDTV--LQSITITLANLT 244
Qy 174 NHMSTNVS-----MTNTLSSMLKGFAPAAACAVQTAQNGVRAMSSL-GSSLS-S 226
Db 245 GPYSIIGLGAIPGGWMLTFGQILGAQNAQVAAALQPKAAAGALSPLAPLRGGYIGDIT 304
Qy 227 GLGGVAAANLGRAASV 242
Db 305 PLGGATGGAIAIVV 320

RESULT 9
F70560
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70560
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: F70560
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-413 <COL>
A: Cross-references: GB:Z95436; GB:AL123456; NID:93261770; PIDN:CAB08826.1; PID:e316565;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 30.9%; Score 403; DB 2; Length 413;
Best Local Similarity 36.9%; Pred. No. 2.7e-20;
Matches 103; Conservative 38; Mismatches 86; Indels 52; Gaps 5;
Db 1 VAMSVTAQAGBELTAQVRAAAAYETAYGLTVPVPIAENRAELMILIAITNLGQNTPA 60
74 VAWLSTAGAEQAGQAQAAASAYEAAFFATVPPVPIAENRALLMALLATNIFLQNTAA 133
Qy 61 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPPEEAPMTSAGGLLEQAAVEEASD 120
Db 134 IAAEAEQAEWMAQDAAMFGVAAATATATATLTPPEEAPMTSAGGLLEQAAVEEASD 192
Qy 121 TAAANQLMNVNPOALKQAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSANMHMNTN 180
Db 193 GAANAQAALTDIPKAL-----FGLSGIFTNPPPLWTLTGKALGLTGTWSSDG 239
Qy 181 SGVSNMTNLTSSMLKGFAPAA--AAQAVOT----- 207
Db 240 SGLIVGGVGLDFVQGVQTSAGELDSVAMDTGKVVSPARLMTVQFQYFGLAHLDPKWS 299

Qy 208 ----AAQNGVRAMSLGSSLSGSGGGVAAANLGRAASV 242
Db 300 EGAKAAGEAKALPAAVPAIPASAGL-SGVAGVGAASV 337

RESULT 10
B70625
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70625
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: B70625
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-391 <COL>
A: Cross-references: GB:Z92539; GB:AL123456; NID:93261714; PIDN:CAB06873.1; PID:e304546;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 30.8%; Score 402; DB 2; Length 391;
Best Local Similarity 39.6%; Pred. No. 2.9e-20;
Matches 99; Conservative 40; Mismatches 99; Indels 12; Gaps 5;
Db 1 VAMSVTAQAGBELTAQVRAAAAYETAYGLTVPVPIAENRAELMILIAITNLGQNTPA 60
73 LAWLTYTAEAAAHAGSQAMASAAAYEAAVMTVPVVAANRALLAALVAINVLGINTPA 132
Qy 61 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPPEEAPMTSAGGLLEQAAVEEASD 120
Db 133 IAAEAEQAEWMAQDAAMFGVAAATATATATLTPPEEAPMTSAGGLLEQAAVEEASD 191
Qy 121 TAAANQ-----LMNVNPOALKQAQPTQGTTPSSKLGGLWKTVPSPHRSPI--ISNMVSAN 173
Db 192 TAAVQVSVADLITSLPNAVSLASFVTSVLDSTGLSGIIADIDALLATPFFVANIINSV 251
Qy 174 NHMS-MTNSGVSMNTLSSMLKGFAPAAACAVQTAQNGVRAMSSLGSSLSGSLGGV 232
Db 252 NTAAVTVNAAITATLFLANLNSGAPVIAEGAIKAEAG--AAASAAAGLADSVTPAGL 308
Qy 233 AANLGRAASV 242
Db 309 GASLGEATLV 318

RESULT 11
C70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: C70931
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-463 <COL>
A: Cross-references: GB:AL0202021; GB:AL123456; NID:93250699; PIDN:CAAL17723.1; PID:e125461;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match	30.8%;	Score 402;	DB 2;	Length 463;
Best Local Similarity	35.7%;	Pred. No. 3.6e-20;		
Matches 114;	Conservative 37;	Mismatches 106;	Indels 62;	Gaps 8;
Qy	1	VAMSVTAGQAEELTAAGVRYVAAAAYETAYGTVTPPVIAENRAELMLIIATNLLGQTPA	60	
Db	72	VAMSAATLAAREAAQASAAAAAYEAAATATVPVVAANRAELAVLAATNIGQNTGA	13-	
Qy	61	IAYNEASYGEMWQDAAMFCYAAATATATATATLLPFEEAPEMTSAGLLEQAAAAVEASD	120	
Db	132	IAAAEARYAEMWQDAAMCYAGSSSVAT-QVTPFPAAPPITTAAGLATQGVAVQAQVG	190	
Qy	121	TAAANQLMNNVPQALKOLAQDTQGTTPSSKUGLWKTVS--PHRSPI-----SN	167	
Db	191	ASAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLMNAVTVGPYASVYVNSMLGLGF AESK	247	
Qy	168	MVSMANNHM-----SMTNSGVSMNTLSSWLKGFAPAA	200	
Db	248	MVLPAQDVTIVSTIFGMVQFQKFFNPVTFENPDLLPKSALGAGLGRSAISSGLGSTAPAI	307	
Qy	201	AAQAVQTAQAQNGVR-----AMBSLSSGLSSGIGGVAANLGRA-----ASV	242	
Db	308	SAGASAGSVGGMSPVPSWAAATPAIRTVAAVFSTGLQAVPAAAISEGSLLSQWALLASV	367	
Qy	243	RYHRDGGKYAXSGRRNGG	261	
Db	368	AGGALGGAAARATGGFTGG	386	

RESULT 12
G70925
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70925
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:9825987; PMID:9634230
A:Accession: G70925
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-408 <COL>
A:Cross-references: GB:274024; GB:AL123456; NID:G3250700; PID:CAA98377.1; PID:c1301025;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

```

Query Match      30.2%; Score 395; DB 2; Length 408;
Best Local Similarity 38.9%; Pred.No.9.2e-20;
Matches 110; Conservative 31; Mismatches 78; Indels 64; Gaps 11;

QY    1 VAMSVTAGOAELTAAQVRVAAAAYETAGTTPPPVAENRAELMILIAINLLGONTPA 60
      |||..|||..|||..|||..|||..|||..|||..|||..|||..|||..|||..
Db     73 VAWLSATAGDAEQMGQARAAAAAYELAFANTVPPVVVAERALLVALVATNFGONTPA 132

QY    61 IAVNEASYGENWAOADAAAMFCYAAATATATATLTPPEAPEMTSAGGLLEGA---V 115
      |||..|||..|||..|||..|||..|||..|||..|||..|||..|||..|||..
Db     133 IAATEAQYAEMWAODAAAMYAGSAAIAT-ELTFPTAAPTTSFPAALAGOAATAVSSTV 191

QY    116 EEASDTAAANCLMN-----VP--QAQKQ-LAOPTOGTTSPSKL-----GGL 154
      |||..|||..|||..|||..|||..|||..|||..|||..|||..|||..|||..
Db     192 PPLATTAAVQLLOQLGSTSLIPHYSAQQWLAEMLLGLPEDNMTIVRLIGISFYDEGL 251

QY    155 WKTVSHRSPISNNMVSWANNHMSVTNGSVNTLTSSMLKGFAFAAAQAQVTAQAQNV 214
      |||..|||..|||..|||..|||..|||..|||..|||..|||..|||..|||..
Db     252 -----LQFEASTLAQQAPCTPGCAG--DSGSVLDSWGPTIFA-----GPR 290

QY    215 AMSSLI--GSSLGSGS-----SGLGGGVAAANLGRAAS 241

```

Db 291 ASPSVAGGAVGVGTPOFYWYWDRESIGSVSAAALKGSS 333

RESULT 13

G70881

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: G70881

R/Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998

A/Authors: Seares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B. A.; Little: Deciphering the biology of Mycobacterium tuberculosis from the genome. Nature 393, 537-544, 1998

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: G70881

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-394 <COL>

A/cross-references: GB:AL008967; GB:AL123456; NTD:G3261491; PIDN:CAA15564.1

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: PPE

Query Match	29.8%;	Score	389;	DB 2;	Length	394;			
Best Local Similarity	37.8%;	Pred.	No. 2.3e-19;						
Matches	103;	Conservative	43;	Mismatches	106;	Indels	22;	Gaps	7;
Qy	1	VAMSVTAGQAEITAAQVRVAAAYATYGLTVPVPVIAENRAELMILIAIATNLLGQNTPA	60						
Db	73	LAWLTVTAESAAHAAQAAMASAAAFEAAPAMTVPPAEVAANRALLAALVAATNVLGQNTPA	132						
Qy	61	IAVNEAEYGEWAQDAAMFGVYAAATATATATILPPEEAPEMTSAGLLLEQAAAVEASD	120						
Db	133	IMATEAHYGEWAQDALAMGYAASAA-AAGRNLNLIITPSQTANMAGLACQAAAVSHAA	191						
Qy	121	TAAANQ-----LMNNYPOALKQDAQPTQGTTPSSKLGGLWKTVSP--HRSPISNMYSMAN	173						
Db	192	ASTVQVQGLGLSLISLNPNAVMGFASPLTSAADAAGLGGIIQDIEELLGITFVQNAINGAV	251						
Qy	174	NHMSMTNSGVSMNTLTSSMLKGFAPAAAAQAVOTAAQNGV---RANSSSLGSSLGSSGLGG	230						
Db	252	N-----TATFVFMATINAVFLGHAFALNPATVTAAADAVPAAAAAAGLAHTTTPFVGVG	307						
Qy	231	-GVAANLGRAASVRYGHRDGGKYAXSGRRNGGPA	263						
Db	308	ASLTASILGEASSV-----GGLSVPAQWSTAAFA	335						

RESULT 14

B70932
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: B70932
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris,
J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares,
N.; Tait, R.; Whitehead, S.; Young, P.; Zeng, Q.; Zeng, X.; Zhou, J.;
Nature 393, 537-544, 1998
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.
A;reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70932
A;status: Preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-468 <CD>
A;cross-references: GB:AL123456; NID:g3250699; PIDN:CAA17730.1
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Job time : 8.35842 secs

[illegible]

RESULT 15

A70646
probable pPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70646
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Connor, R.; Davies, R.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Seares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70646
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <COL>
A:Cross-references: GB:I283867; GB:A123456; NID:G3261695; PID:CAB06278.1; PID:G291015;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: pPE

```

Query Match      29.0%; Score 379; DB 2; Length 380;
Best Local Similarity 36.0%; Pred. No. 1e-18;
Matches 108; Conservative 40; Mismatches 96; Indels 56; Gaps 10;

QY      1 VAWMSVTAGBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIATNLLGONTPA 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      73 IGLVLTAEKTCQTATQARAALAFQRAYAMTLPFPVVAANRIQLLALATNFFQNTAA 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      61 IAVNEAYGEKWAQDAAMAFYAAATATATATLTPPEAPEMTSAGGLLEQAAAVVEASD 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      133 IAATEAQVAEKWAQDAAMAYGYATASA-AAALLTPFPSPQTTNPAGLTAQAAAVSQATD 191
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      121 TAAANQLMKNYPQALKQLAQP-----TQGT-----PSK 150
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      192 PLSL--LIETVQALQALTI SPIDFTFLDAIFAGYATGVTDVDSFVAGTICAESN 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      151 LGL-----WKTYSPHRSPISNNVSMANNHMTNSG---VSWNTLLSML----- 193
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      250 LGLLVNGDENPAEVTGDFGELVLTSPCGGVSASAGGAASVNGVTIASVGRANISG 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      194 -KGFPAAAAQAVQTAAQNGVRAMSSL-GSSLGSSLGG--GYAANLGRAASV--RYGHR 247
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      310 QLSVPPSNAAPSTRPVSAISPAGLTILPGTDVAEHGMPGVPVPVGAASGLVPRYGV 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: June 22, 2004, 17:24:51

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 4.37699 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-12

Perfect score: 1306

Sequence: 1 VAMSVTAQGAELTAQVRV.....YGHRDGKXVXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1001.5	76.7	396	1 YD61_MYCTU	Q11031 mycobacteri
2	402	30.8	463	1 Y102_MYCTU	O53951 mycobacteri
3	395	30.2	408	1 YS92_MYCTU	Q10813 mycobacteri
4	255.5	19.6	678	1 YF48_MYCTU	Q10778 mycobacteri
5	243.5	18.6	443	1 Y878_MYCTU	Q10540 mycobacteri
6	243	18.6	487	1 Y442_MYCTU	P42611 mycobacteri
7	194.5	14.9	463	1 Y096_MYCTU	Q10892 mycobacteri
8	191.5	14.7	408	1 SRA_MYCLE	Q07297 mycobacteri
9	158.5	12.1	434	1 YU18_MYCTU	P31500 mycobacteri
10	158	12.1	435	1 YU22_MYCTU	O53268 mycobacteri
11	156	11.9	178	1 Y129_MYCTU	O06246 mycobacteri
12	152	11.6	176	1 Y125_MYCTU	O50703 mycobacteri
13	114.5	8.8	1211	1 BUN2_DROME	Q24523 drosophila
14	112	8.6	2090	1 N214_HUMAN	P35658 homo sapien
15	110.5	8.5	938	1 TRG1_ECOLI	P33790 escherichia
16	107.5	8.2	354	1 YAU6_SCHPO	Q10169 schizosacch
17	106	8.1	1783	1 RAA3_CHLRE	Q3f6c4 chlamydomon
18	103.5	7.9	881	1 PRY3_YEAST	P47803 saccharomyc
19	102	7.8	1120	1 STFR_ECOLI	F76072 escherichia
20	100.5	7.7	1537	1 FLO1_YEAST	P32768 saccharomyc
21	99	7.6	329	1 PRY2_YEAST	P36110 saccharomyc
22	99	7.6	659	1 SFRG_HUMAN	O8n2m8 homo sapien
23	99	7.6	1140	1 YM96_YEAST	Q04993 saccharomyc
24	98.5	7.5	475	1 SIM1_YEAST	P40472 saccharomyc
25	98.5	7.5	553	1 FXC1_MOUSE	Q61572 mus musculu
26	98.5	7.5	558	1 YJ83_MYCTU	Q10873 mycobacteri
27	98.5	7.5	577	1 CST2_HUMAN	P33240 homo sapien
28	98.5	7.5	743	1 PO21_HUMAN	P14859 homo sapien
29	98.5	7.5	1075	1 FLO5_YEAST	P38894 saccharomyc
30	98.5	7.5	1845	1 Z236_HUMAN	O9u136 homo sapien
31	98	7.5	825	1 ICP0_HSV2H	F28284 herpes simp
32	97.5	7.5	416	1 CREA_EMENI	Q01981 emericella
33	97.5	7.5	526	1 NU62_MOUSE	Q63850 mus musculu

34 96.5 7.4 461 1 US45_LAGLC
35 96.5 7.4 461 1 Y514_HUMAN
36 96.5 7.4 632 1 PO21_RAT
37 96 7.4 939 1 SLAP_CAMPE
38 96 7.4 1025 1 SLAP_CAUCR
39 95.5 7.3 2035 1 HFC1_HUMAN
40 95.5 7.3 2090 1 HFC1_MESAU
41 95 7.3 323 1 JUND_CHICK
42 95 7.3 774 1 STF_IAMD
43 95 7.3 1556 1 PRO3_DROVI
44 94.5 7.2 601 1 PM1_DROME
45 94.5 7.2 780 1 CTPA_MYCLE

ALIGNMENTS

RESULT 1

YD61_MYCTU
ID YD61_MYCTU STANDARD; PRT; 396 AA.
AC Q11031;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rvi361c/MT1406.
GN Rvi361c OR MT1406 OR MTCY02B10.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagoe K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains".
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; 275555; CAA99966.1;
DR EMBL; AE007013; AAK45669.1; --
PIR; H70741; H70741.
DR TIGR; MT1406;
DR Tuberculist; Rvi361c;
DR InterPro; IPR000030; Microbac_PPE.

DR Pfam: PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 158 159 TA -> AT (IN REF. 2).
 SQ SEQUENCE 396 AA; 40015 MW; 6AF6AE0D7B5F668D0 CRC64;

Query Match 76.7%; Score 1001.5; DB 1; Length 396;
 Best Local Similarity 84.1%; Pred. No. 2.4e-62;
 Matches 207; Conservative 13; Mismatches 21; Indels 5; Gaps 2;

QY 1 VAMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 74 VAMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 133
 QY 61 IAVNEAEYGENWAQDAAMFGYAATAATATATLLPPEEAPEMTSAGLLLEQAAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAATAATATATLLPPEEAPEMTSAGLLLEQAAAVEASD 193
 QY 121 TAAANQLMNNVPQALKQIAQPTQGTPTSSKLGGLKWTVPSPHSPISNMVSMANNHSMTN 180
 DB 194 TAAANQLMNNVPQALKQIAQPTQGTPTSSKLGGLKWTVPSPHSPISNMVSMANNHSMTN 253
 QY 181 SGVSMNTLSMLKGFAPAAAQAVTAQAQNGVPMSS----LGSSLSGSLGGGVAANL 236
 DB 254 SGVSMNTLSMLKGFAPAAAQAVTAQAQNGVPMSS----LGSSLSGSLGGGVAANL 312
 QY 237 GRAASV 242
 DB 313 GRAASV 318

RESULT 2
 Y102 MYCTU STANDARD; PRT; 463 AA.
 AC O53951;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv1802/MT1851/Mb1830.
 GN Rv1802 OR MT1851 OR MT049.24 OR Mb1830.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium tuberculosis, and
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL J. Bacteriol. 184:5479-5490 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.

RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garner T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RA "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL022021; CAA17723.1; -;
 DR EMBL; AE007044; AAK46123.1; -;
 DR EMBL; BX248340; CAD94533.1; -;
 DR EIR; C70931; C70931.
 DR TIGR; MT1851; -;
 DR TubercuList; Rv1802; -;
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 401 401 S -> L (IN REF. 2).
 FT SEQUENCE 463 AA; 46021 MW; E564828BF09FA551 CRC64;

Query Match 30.8%; Score 402; DB 1; Length 463;
 Best Local Similarity 35.7%; Pred. No. 8.2e-21;
 Matches 114; Conservative 37; Mismatches 106; Indels 62; Gaps 8;

QY 1 VAMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 72 VAMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTGA 131
 QY 61 IAVNEAEYGENWAQDAAMFGYAATAATATATLLPPEEAPEMTSAGLLLEQAAAVEASD 120
 DB 132 IAAEAARYAEMWAQDAAMFGYAATAATATATLLPPEEAPEMTSAGLLLEQAAAVEASD 190
 QY 121 TAAANQLMNNVPQALKQIAQPTQGTPTSSKLGGLKWTVPSPHSPISNMVSMANNHSMTN 167
 DB 191 ASAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLNAVTVGPYASSVYNSMGLGFAESK 247
 QY 168 MYSMANNH--
 DB 248 MYLPANDTVISTFGVQFQFFNPVTFPNDLPKSAAGAGLGLRGAISSGLGTAPAI 307
 QY 201 AAQAVTAQAQNGVR-----AMSLGSSLSGSLGGGVAANLGEA-----ASV 242
 DB 308 SAGASQAGSVGSGMSPPSMAAATPAIRTVAAVFSTGLQAVPAAAISEGSLLSQMALASV 367
 QY 243 RYGHDRDGKXVXSGRRNGG 261
 DB 368 AGGALGGAAARATGGFLGG 386

RESULT 3
 YS92 MYCTU STANDARD; PRT; 408 AA.
 AC Q10513;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv2892c/MT2959/Mb2916c.
 GN Rv2892c OR MT2959 OR MTCY274.23C OR Mb2916c.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;

```

RN RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moulé S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Kolonay J.F., Nelson W.C., Weidman J., Khouri H., Gill J., Mikula A.,
RA Delcher A., Utterback T., Weidman J., Venter J.C., Fraser C.M.;
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigemeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
RL [1]
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z74024; CA98377.1; --
CC EMBL; AB007119; AAK47285.1; --
CC EMBL; BX248344; CAD96603.1; --
CC PIR; G70925; G70925.
CC TIGR; MT2959; --
CC Tuberculist; Rv2892c; --
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
CC Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 56 76 POTENTIAL.
SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F20D7827199 CRC64;

Query Match 30.2%; Score 395; DB 1; Length 408;
Best Local Similarity 38.9%; Pred. No. 2.2e-20;
Matches 110; Conservative 31; Mismatches 78; Indels 64; Gaps 11;

QY 1 VAMSVTAGQAELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAITNLGQNTPA 60
DB 73 VAMLSATAGQAEGMQARAAAAYELAFAMTVPPVIVANRALLVALVATNFEQNTPA 132
QY 61 IAVNEAEYGEWQAQDAAMFYAAATATATATLTPFEAPEMTAGGLLEQAAA-----V 115
DB 133 IAAATEAQYAEWQAQDAAMFYAAGSAATAT-ELTFFTAAPVTTSPAALAGQAATVSTV 191
QY 116 EASDSTAANQIMNN-----VP--QALKQ-LAQPTQCTTPSSKL-----GGL 154

```

```

DB 192 PPLATTAAPQLQSLSTSLIPWYSALQQLWAENLLGLTPDNRMNTIVRLIGISYFDEGL 251
QY 155 WKTVPSPHRSPISNMYSVANNHMSMTNSGVSNWTLSSMLKGFAPAAQAQVQTAQNGVR 214
DB 252 -----LQFEASLAQAIPGTPGAG--DGSVSLDSWGPTIFA-----GPR 290
QY 215 AMSSL--GSSLGS-----SGLGCGGVAANLGRAAS 241
DB 291 ASPSVAGGAGVGQVTPQPYWYMWALDRESIGSVSAALGKSS 333
RESULT 4
YF48 MYCTU
ID YF48 MYCTU STANDARD; PRT; 678 AA.
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1548c/MT1599.
GN Rv1548C OR MT1599 OR MTCY48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moulé S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Kolonay J.F., Nelson W.C., Weidman J., Khouri H., Gill J., Mikula A.,
RA Delcher A., Utterback T., Weidman J., Venter J.C., Fraser C.M.;
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z74020; CA98335.1; --
CC EMBL; AB007026; AAK45866.1; ALT_INIT.
CC PIR; A70762; A70762.
CC TIGR; MT1599; --
CC Tuberculist; Rv1548c; --
CC InterPro; IPR000030; Microbac_PPE.
CC InterPro; IPR002989; Mycobac_pentapep.
CC Pfam; PF01469; Pentapeptide_2; 11.
CC Pfam; PF00823; PPE; 1.
CC Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 180 200

```



```

RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15467; AAA88235.1; ALT INIT.
DR EMBL; AL021932; CAA17399.1; -
DR EMBL; AF006948; AAK44681.1; -
DR PIR; C70830; C70830.
DR TIGR; MT0458; -
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 5.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 40 40 E -> K (IN REF. 2).
FT CONFLICT 96 96 I -> T (IN REF. 1).
FT CONFLICT 211 211 G -> GNNIG (IN REF. 1).
SQ SEQUENCE 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;

Query Match
Best Local Similarity 18.68; Score 243; DB 1; Length 487;
Matches 83; Conservative 30; Mismatches 129; Indels 46; Gaps 8;

QY 1 VAMSVTAGAQLTAQVRAAAAYETAYGLTVPVPIAENRAELMILIANLLGNTPA 60
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 LAMLSIAAQAEQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 135
QY 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLPPFEAPEMTSAG-----GLLEQAAA 114
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 LMDVEAAAYEQWALDVAAGYHFDASAAVAQAAPQOV--LRNLGIDIGKNGQINLFG 193
QY 115 VEASDTPAANQLMNNVPQALKQAQPTQGTTPSKLGL-----GLKKTVS----- 159
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 NTGSGNTGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNG 253
QY 160 ---PHRSPFINNVSWANNHMTSMVNTNTLSMLKGFAPAAAQAQVTAQAQNGVRAM 216
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ITGDHQMGFGFGNFGSGGN-IGFGNFGTGNVGLFNS-----GSGNIGIGNS 297
QY 217 SSLGSSLLGSSGLGGVAANLGRASVRYGHRDGG-KVAXSGRRNGGPA 263
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 GSLNSGIGTSGT---INAGLSAGSLNTSFNAGNQNALGSAAGSEA 342

```

```

RESULT 7
Y096 MYCTU STANDARD; PRT; 463 AA.
ID Q10892;
AC Q10892;1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0096/MT0105.
GN RV0096 OR MT0105 OR MTCY251.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74410; CAA98932.1; -
DR EMBL; AF006922; AAK44327.1; -
DR PIR; H70750; H70750.
DR TIGR; MT0105; -
DR TubercuList; RV0096; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
SQ SEQUENCE 463 AA; 46894 MW; 42D9D66A033D0DD8 CRC64;

Query Match
Best Local Similarity 14.9%; Score 194.5; DB 1; Length 463;
Similarity 36.3%; Pred. No. 1.8e-06;

```

Matches 53; Conservative 21; Mismatches 55; Indels 17; Gaps 3;
 QY 1 VAWMSVTAGQAEHTAAQVRAAAAYETAYGLTVPPVIAENRAELMILIAATNLGQNTTPA 60
 DB 70 LAWEQTAINSAAVTAQHQHAAAYCSALAAMPAPAEALAAHAIHGVLIATNFFGINTVP 129
 QY 61 IAVNEAEYCEMWAQDAAMFGVAAATATATATLLPFEEAPEMTSAGGLLEQAAVEASD 120
 DB 130 IALNEADYVWQLQADTWAYQAVADATVAVSTQPPAPIRAPGG-----DAAD 180
 QY 121 T-----AAANQMLNVPQALKQLAQP 141
 DB 181 TRLDVLSSIGQLIRDI---LDFIANP 203

RESULT 8

SRA_MYCLE SRA_MYCLE STANDARD; PRT; 408 AA.
 AC Q07297;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine-rich antigen (25k) (45 kDa protein).
 GN SRA OR ML0411 OR MLC1383.14.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9329328; PubMed=8478104;
 RA Vega-Lopez F., Brooks L.A., Dockrell H.M., de Smet K.A.L.,
 RA Thompson J.K., Hussain R., Stoker N.G.;
 RT "Sequence and immunological characterization of a serine-rich antigen
 from Mycobacterium leprae";
 RL Infect. Immun. 61:2145-2153(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9502054; PubMed=7934845;
 RA Rinke de Wit T.F., Clark-Curtiss J.E., Abebe F., Kolkon A.H.J.,
 RA Jonson A.A.M., Thole J.E.R.;
 RT "A Mycobacterium leprae-specific gene encoding an immunologically
 recognized 45 kDa protein";
 RL Mol. Microbiol. 10:829-838(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 RL -1- SIMILARITY: Belongs to the mycobacterial PPE family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00015; AAC43220.1; -
 CC EMBL; X69431; CAA48480.1; -
 CC EMBL; Z21952; CAA79950.1; -
 CC EMBL; Z97179; CAB09938.1; -
 CC RP

DR EMBL; AL583918; CAC29919.1; -.
 DR PIR; C86960; C86960.
 DR PIR; S33522; S33522.
 DR PIR; S39872; S39872.
 DR Lepidoma; ML0411; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Antigen; Repeat; Complete proteome.
 FT DOMAIN 192 196 POLY-SER.
 FT REPEAT 209 235 2 X 6 AA REPEATS OF S-V-A-Q-S-E.
 FT REPEAT 230 235 1.
 FT CONFLICT 132 132 T -> S (IN REF. 2).
 FT CONFLICT 189 189 S -> L (IN REF. 2).
 FT CONFLICT 191 191 H -> D (IN REF. 2).
 FT CONFLICT 292 292 P -> L (IN REF. 2).
 SQ SEQUENCE 408 AA; 42466 MW; 5C0C2BE0D6E6A9D8 CRC64;
 Query Match 14.7%; Score 191.5; DB 1; Length 408;
 Best Local Similarity 24.5%; Pred. No. 2.6e-06;
 Matches 70; Conservative 49; Mismatches 104; Indels 63; Gaps 8;
 QY 1 VAWMSVTAGQAEHTAAQVRAAAAYETAYGLTVPPVIAENRAELMILIAATNLGQNTTPA 60
 DB 74 VAWLDGNAENAGLIAARVLHVAVAFAEAPAGVMVPLLTVLGNIHTMALKA-NWFGQVSTT 133
 QY 61 IAVNEAEYCEMWAQDAAMFGVAAATATATATLLPFEEAPEMTSAGGLLEQAAVEASD 120
 DB 134 VAALEADYDLMMVQNSTAMTTTRDTVLRETGMENFEPAPOLVSR-YCMDRDSVNSFHS 192
 QY 121 TAAANQMLNVPQALKQLAQP-----PTQ----- 143
 DB 193 SSSSDSLVESIDNLYDSVAQSEEHGSDMSQSNTCGSVAQSELCDSPFGPSSQSSND 252
 QY 144 -GTTPSKLGGLWKTSPHRSPISNMVSMANNHMTNNTLSSMLKGFAPAAA 201
 DB 253 LSATSLTQQLGGL-----DSIISSASALLTTNS--ISSSTASSIM---PIVA 295
 QY 202 AQAVTAAQNGVRAMSSLSGLSGGGVAAAN-----LGRAASV 242
 DB 296 SQVETLGRSQV-AVERMIQSISSSTAVSDVVAASKVAGVQCAVSV 340

RESULT 9

YU18 MYCTU YU18 MYCTU STANDARD; PRT; 434 AA.
 AC P31500; OS1265;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV3018C/MT3098/MT3101.
 GN RV3018C OR MT3098/MT3101 OR MTV012.32C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., Maclean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R., Squares J.,
 RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

DR InterPro: IPR000030; Microbac_PPE.
DR Pfam: PF00823; PPE; 1.
DR Hypochemical protein; Complete proteome.
SQ SEQUENCE 178 AA; 19811 MW; 8BEIFC025ABFBEA6 CRC64;

Query Match 11.9%; Score 156; DB 1; Length 178;
Best Local Similarity 37.1%; Pred. No. 0.00029;
Matches 39; Conservative 16; Mismatches 50; Indels 0; Gaps 0;

QY 3 WMSVTAGQAELTAQVRVAAAAYETAYGLTVPPPIAENRAELMLLIATNLGQNTPAIA 62
Db ::::|
73 WLKHSRQLLETAYVIDFLAYVEETHRVKVVPPTAIANREHRLIASNVAGVNTPAIA 132
QY 63 VNEAEYGEMWAQDAAMFCYAATAATATATLTPFEAPEMTSAG 107
Db ::::|
133 GLDAQIQQTFRAGNIWVDYQSTARFILAYLPRWQEPPOIYGGG 177

RESULT 12
YY25 MYCTU
ID YY25 MYCTU STANDARD; PRT; 176 AA.
AC Q50703;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein Rv3425.
GN RV3425 OR MTCV78_04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetaceae;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.F., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.B., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licens@isb-sib.ch).

EMBL; Z77165; CABO1031.1; -.
DR PIR; F70738; F70738.
DR Tuberculist; RV3425; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 176 AA; 19855 NW; BCCEP2E9463B87B0 CRC64;

Query Match 11.6%; Score 152; DB 1; Length 176;
Best Local Similarity 36.5%; Pred. No. 0.00054;
Matches 38; Conservative 17; Mismatches 49; Indels 0; Gaps 0;

QY 3 WMSVTAGQAELTAQVRVAAAAYETAYGLTVPPPIAENRAELMLLIATNLGQNTPAIA 62
Db ::::|
73 WLKSHSGTKHAAMVINGLANAYNDTRKKVVPPPEIAANREERRLIASNVAGVNTPAIA 132
QY 63 VNEAEYGEMWAQDAAMFGYAAATATATATLTPFEAPEMTSAG 106

```

DB 133 DUDAQDYQYRARNVAVMNAVYVWTRSAISDLPRWREPPPIYRG 176
RESULT 13
BUN2_DROME
ID_BUN2_DROME STANDARD; PRT: 1211 AA.
AC Q24523; Q9VK78; Q9VK79;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bunched protein, class 2/class 3 isoforms (Shortsighted protein).
GN BUN OR SHS OR CG5461
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1] _SEQUENCE FROM N.A. (ISOFORM CLASS 2), SUBCELLULAR LOCATION, AND
RP FUNCTION.
RC TISSUE=Eye-antennal disk;
RX MEDLINE=96038094; PubMed=7555710;
RA Treisman J.E., Lai Z.-C., Rubin G.M.;
RT "Shortsighted acts in the decapentaplegic pathway in Drosophila eye
development and has homology to a mouse TGF-beta-responsive gene.";
RL Development 121:2835-2845(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS CLASS 2 AND CLASS 3).
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champ M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.B.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berwan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Buck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Snie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "the genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: Probable transcription factor required for peripheral
CC nervous system morphogenesis, eye development and oogenesis. May
CC be required for the transmission of the dpp signal and for a
CC morphogenetic movement of the medulla in the brain that recruits

```

```

CC the second optic lobe relative to the first. plays a role in
CC determining proper dorsal cell fates leading to the formation of
CC the dorsal appendages.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC NamesClass 2;
CC IsoId=Q24523-1; Sequence=Displayed;
CC NamesClass 1;
CC IsoId=Q24522-1; Sequence=External;
CC NamesClass 3;
CC IsoId=Q24523-2; Sequence=VSP 006670;
CC -!- SIMILARITY: Belongs to the TSC-22/Dip/Bun family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L42512; AAC41608.1; ..
CC EMBL; AE003636; AAF53200.1; ALT_SEQ.
CC EMBL; AE003636; AAF53201.1; ALT_SEQ.
CC PIR; T13804; T13804.
CC HSP; P80220; 1DIP.
CC FlyBase; FBgn0010460; bun.
CC GO; GO:0007432; P:peripheral nervous system development; NAS.
CC InterPro; IPR000580; TSC-22_Dip_Bun.
CC Pfam; PF01166; TSC22; 1.
CC ProDom; PD007152; TSC-22_Dip_Bun; 1.
CC PROSITE; PS01289; TSC22; 1.
CC Transcription regulation; Nuclear protein; Alternative splicing.
CC DOMAIN 15 31
CC DOMAIN 76 86
CC POLY-GLN.
CC POLY-GLN.
CC POLY-GLN.
CC POLY-SER.
CC POLY-SER.
CC POLY-SER.
CC POLY-HIS.
CC POLY-GLY.
CC POLY-GLY.
CC POLY-GLN.
CC POLY-GLN.
CC POLY-ALA.
CC POLY-ALA.
CC POLY-ALA.
CC POLY-ALA.
CC GLN-RICH.
CC POLY-ALA.
CC POLY-ALA.
CC LEUCINE-ZIPPER.
CC POLY-ALA.
CC Missing (in isoform Class 3).
CC /Ftid=VSP 006670.
CC K -> E (IN REF. 1).
CC MISSING (IN REF. 1).
CC Q -> CQQ (IN REF. 1).
CC QVTSAA -> TS (IN REF. 2).
CC SEQUENCE 1211 AA; 125299 MW; 9925A9159A7051B0 CRC64;
SQ
Query Match 8.8; Score 114.5; DB 1; Length 1211;
Best Local Similarity 25.9%; Pred. No.1.8;
Matches 66; Conservative 28; Mismatches 80; Indels 81; Gaps 13;
QY 14 TAAQVRVAAAYETAYGLTPVPVIAENRAELMILIAITNLGQNTFAIA---VNEAEYGE 70

```

Db 828 TAQQ---AAAAAATSAVTPPPQNTSNAAV---TTGGQQTMLLSHMTSVEQQQPN 881
 QY 71 MWAQDAAMFYAAATATATATLIP---FEAPPMTSAGLLEQAABEASDTAAANQL 127
 Db 882 LGAAAAAAGGAATSAVAPQAIPTLQLQSPAP-----STIADPQQL 923
 QY 128 MNVPFQAL-----KOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMWSMA--NNHMS 178
 Db 924 M--VPOQQQQQHOEQQQQPPQQQQP-----LPPANIASANSNINL 966
 QY 179 TN-----SGVMTNTL-----SSMLKGPAPAAAQAV-QTAA-----QNGV 213
 Db 967 TINTVATGEATNTALTITDEQATAALAAAFATGAAAAATGATSAATAATQCOIQLOQOP 1026
 QY 214 RAMSSSLGSSGL 228
 Db 1027 NAESETESAGTSV 1041

RESULT 14

N214 HUMAN
 ID N214 HUMAN STANDARD; PRT; 2090 AA.
 AC P35658;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Nuclear pore complex protein Nup214 (Nucleoporin Nup214) (214 kDa nucleoporin) (CAN protein).
 DB NUP214 OR CAN OR CAIN.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=92195315; PubMed=1549122;
 RA Von Lindern M., Fornerod M., Van Baal S., Jaegle M., De Wit T., Buijs A., Grosveld G.;
 RT "The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA."; Mol. Cell. Biol. 12:1667-1697(1992).
 RL [2]
 RN CHARACTERIZATION.
 RX MEDLINE=94151361; PubMed=9108440;
 RA Kraemer D., Wozniak R.W., Blobel G., Radu A.;
 RT "The human CAN protein, a putative oncogene product associated with myeloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm."; Proc. Natl. Acad. Sci. U.S.A. 91:1519-1523(1994).
 RL
 CC -!- FUNCTION: May serve as a docking site in the receptor-mediated import of substrates across the nuclear pore complex.
 CC -!- SUBUNIT: Homodimer. Interacts with DDX19 and NUP88.
 CC -!- SUBCELLULAR LOCATION: Nuclear pore complex. Cytoplasmic filaments.
 CC -!- TISSUE SPECIFICITY: Expressed in thymus, spleen, bone marrow, kidney, brain and testis, but hardly in all other tissues or in whole embryos during development.
 CC -!- DOMAIN: Contains F-G repeats.
 CC -!- PTM: Probably glycosylated as it reacts with wheat germ agglutinin (WGA).
 CC -!- DISEASE: Implicated in a subset of acute myeloid leukemia (acute nonlymphocytic leukemia) (AML) carrying a chromosomal translocation t(6;9)(p23;q34) that results in the formation of a DEK-CAN fusion gene.
 CC -!- DISEASE: Involved in some cases of acute undifferentiated leukemia (AUL) through a chromosomal translocation t(6;9)(q21;q34.1) that involves NUP214/CAN and SET.
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chronocancer/Genes/CAN.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X64228; CAA45535.1; -.
 DR PIR; S26058; S26058.
 DR Genew; HGNC:8064; NUP214.
 DR MIM; 114350; -.
 DR GO; GO:0005643; C:nuclear pore; TAS.
 DR GO; GO:0005215; F:transporter activity; TAS.
 DR InterPro; IPR004325; Nucleoporin_FG.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF03093; Nucleoporin_FG; 22.
 DR SMART; SM00320; WD40; 2.
 DR Nuclear protein; Transport; Proto-oncogene; Chromosomal translocation;
 KW Repeat; Glycoprotein.
 FT DOMAIN 481 2076 11 X 5 AA APPROXIMATE REPEATS.
 FT DOMAIN 1409 2084 18 X 4 AA APPROXIMATE REPEATS.
 FT DOMAIN 1427 2085 11 X 3 AA APPROXIMATE REPEATS.
 FT DOMAIN 1213 2090 PRO/SER/THR-RICH.
 FT DOMAIN 740 768 LEUCINE-ZIPPER 1.
 FT DOMAIN 861 882 LEUCINE-ZIPPER 2.
 FT SITE 812 813 BREAKPOINT.
 SQ SEQUENCE 2090 AA; 213766 MW; 6DBE767FDD857F8F CRC64;
 Query Match 8.6%; Score 112; DB 1; Length 2090;
 Best Local Similarity 23.1%; Pred No. 5;
 Matches 71; Conservative 37; Mismatches 130; Indels 70; Gaps 13;
 QY 9 QOALTAQAQVRVA-----AAAYET---AYGLTVPPPI 38
 Db 1592 GQTAVTAAGSAGPVAVETSTPIASSTTSIVAPGPSAAAFGTGSSSVPAQPPAA 1651
 QY 39 AENRAELMLLTATNLGONTPTAIAVNEAEYGMQAQDAAMFGVAAATATATLTP--- 95
 Db 1652 SSSS-----AFNQLTNTTATPSATPVFCQVAASTAPSLFGQQTGSTASTAAATQVVS 1704
 QY 96 ---FEAPEMTSAGLLEQA-----AAVEEASDTAAANQLMNVNP--QALKOLAQTOGTT 146
 Db 1705 SSGFSSPAFGTTPGVFCQQTTFGQASVFGQSASSAASVFSQPCFSSVPFAGQPA-SST 1763
 QY 147 PSSKLGGLWKTVPSPHRSPISNMWSMANNHMTNSGVSMNTLSMLKGPAP----- 198
 Db 1764 PTSTSGSVFGAASSTSS--SSSFSG---QSSPNTGGGLFGQSNAPAFGOSPGFGQGSV 1818
 QY 199 -AAAAAQVQTAQNGVRAMSSLSGLSSGLGG--GVAANLGRAASVRYGHRDGGKYAXS 255
 Db 1819 FGGTSAATTTAATSGFSPCQA--SGFGSNTGVSFGQAASGTGGIVFGQQSSSSSGSVFGS 1876
 QY 256 GR--RNGG 261
 Db 1877 GNTGRGGG 1884
 RESULT 15
 TRGL_ECOLI
 ID TRGL_ECOLI STANDARD; PRT; 938 AA.
 AC P33750;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trg protein.
 GN Escherichia coli.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 CC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 21.5043 Seconds
(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349a-12

Perfect score: 1306

Sequence: 1 VAMWSVTAGQAEUAAQVRV.....YGHDDGGKYXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriapi:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1187	90.9	391	16	O05298 mycobacteri
2	1158.5	88.7	390	16	Q7U0E9 mycobacteri
3	1122.5	85.9	396	16	Q7U071 mycobacteri
4	1008.5	77.2	393	16	O8VIZ3 mycobacteri
5	949.5	72.7	393	16	O06341 mycobacteri
6	949.5	72.7	393	16	Q7UWFS mycobacteri
7	463	35.5	399	16	Q7TZH8 mycobacteri
8	460	35.2	403	16	O53956 mycobacteri
9	441	33.8	423	16	O53950 mycobacteri
10	441	33.8	423	16	Q7TZI4 mycobacteri
11	432.5	33.1	393	16	Q7TZJ3 mycobacteri
12	432.5	33.1	411	16	O53939 mycobacteri
13	424.5	32.5	421	16	Q925K0 mycobacteri
14	422	32.3	409	16	O53957 mycobacteri
15	420	32.2	409	16	Q7TZH7 mycobacteri
16	420	32.2	410	2	Q99Q11 mycobacteri

17	403	30.9	413	16	O06386 mycobacteri
18	402	30.8	391	16	P96362 mycobacteri
19	402	30.8	391	16	Q7U075 mycobacteri
20	390	29.9	394	16	Q7TXX5 mycobacteri
21	389	29.8	462	16	O33310 mycobacteri
22	387.5	29.7	694	16	O8VJW0 mycobacteri
23	386.5	29.6	468	16	O53958 mycobacteri
24	379	29.0	380	16	P95190 mycobacteri
25	379	29.0	380	16	Q7TX66 mycobacteri
26	375	28.7	385	16	O33204 mycobacteri
27	375	28.7	385	16	O8VJZ0 mycobacteri
28	375	28.7	385	16	Q7TZR7 mycobacteri
29	361	27.6	381	16	Q7TX67 mycobacteri
30	355	27.2	382	16	Q7TXX3 mycobacteri
31	355	27.2	397	2	Q9AGF0 mycobacteri
32	352	27.0	402	16	O33312 mycobacteri
33	348.5	26.7	365	16	O86373 mycobacteri
34	347.5	26.6	405	16	O8VJW5 mycobacteri
35	347	26.6	364	16	Q7TZJ5 mycobacteri
36	346.5	26.5	394	16	O33205 mycobacteri
37	346.5	26.5	394	16	Q7TZB6 mycobacteri
38	346	26.5	350	16	Q7TZJ2 mycobacteri
39	346	26.5	363	16	O53940 mycobacteri
40	339.5	26.0	443	16	O8VKL9 mycobacteri
41	333.5	25.5	443	16	Q7U242 mycobacteri
42	332.5	25.5	187	16	Q7TZH6 mycobacteri
43	329	25.2	391	16	O05798 mycobacteri
44	329	25.2	391	16	Q7TX76 mycobacteri
45	328	25.1	406	16	P71869 mycobacteri

ALIGNMENTS

RESULT 1

O05298 PRELIMINARY; PRT; 391 AA.

AC O05298;
DT 01-JUL-1997 (TREMELrel. 04, Created)
DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein (PPE family protein).
GN RV1196 OR MTC1364.08 OR MT1234.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F., Davies R., Basham D., Brown D., Chillingworth T., Connor R., Hornsby T., Devlin K., Feltwell T., Gentles S., Hamlin N., Halloway S., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
RL Submitted (APR-2001) to the ENBL/GenBank/DBJ databases.

DR EMBL; Z93777; CAB07839.1; -.
 DR EMBL; AE007000; AAK45491.1; -.
 DR PIR; B70608; B70608.
 DR TIGR; MT1234; -.
 DR TubercuList; Rv1196; -.
 DR InterPro; IPR00030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 391 AA; 39158 MW; E409396B3ABDC0F8 CRC64;

Query Match 90.9%; Score 1187; DB 16; Length 391;
 Best Local Similarity 99.6%; Pred. No. 4.5e-73;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGLLFQAAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLFQAAAVEASD 193

QY 121 TAAANQLMNNVPQALKQIAQPTQGTTPSSKLGGLWKTVPSPHSPTSNMVSMANNHSMTN 180
 DB 194 TAAANQLMNNVPQALKQIAQPTQGTTPSSKLGGLWKTVPSPHSPTSNMVSMANNHSMTN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSLGSSGLGGVAAANLGRAA 240
 DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSLGSSGLGGVAAANLGRAA 312

Query Match 88.7%; Score 1159.5; DB 16; Length 390;
 Best Local Similarity 97.9%; Pred. No. 3.9e-71;
 Matches 237; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLFQAAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLFQAAAVEASD 193

RESULT 2

QYU0E9 PRELIMINARY; PRT; 390 AA.
 AC Q7U0E9;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE18 OR MB1228.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248338; CAD94089.1; -.
 KW Complete proteome.
 SQ SEQUENCE 390 AA; 42788276BAB0B436 CRC64;

RESULT 3

QYU071 PRELIMINARY; PRT; 396 AA.
 AC Q7U071;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE19 OR MB1396C.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248338; CAD94257.1; -.
 KW Complete proteome.
 SQ SEQUENCE 396 AA; 39651 MW; 698A92A50E3CBA3F CRC64;

Query Match

85.9%; Score 1122.5; DB 16; Length 396;
 Best Local Similarity 94.3%; Pred. No. 1.1e-68;
 Matches 232; Conservative 5; Mismatches 4; Indels 5; Gaps 2;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLFQAAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLFQAAAVEASD 193

QY 121 TAAANQLMNNVPQALKQIAQPTQGTTPSSKLGGLWKTVPSPHSPTSNMVSMANNHSMTN 180
 DB 194 TAAANQLMNNVPQALKQIAQPTQGTTPSSKLGGLWKTVPSPHSPTSNMVSMANNHSMTN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSLGSSGLGGVAAANL 236
 DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSLGSSGLGGVAAANL 312

QY 237 GRAASV 242
 DB 313 GRAASV 318

RESULT 4

Q8VIZ3 PRELIMINARY; PRT; 393 AA.
 AC Q8VIZ3;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; Z95390; CAB08702.1; --
DR PIR; C70568; C70568.
DR TubercuList; RV3478; --
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 72.7%; Score 949.5; DB 16; Length 393;
Best Local Similarity 79.7%; Pred. No. 7e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;

QY 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
DB 74 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLEQAAVEEASD 120
DB 134 IEANQAAVSQMWGQDAEAMGYAATAATATATALLPPEADPLITNPGGLLEQAVAVEAID 193
QY 121 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLKWTVPSPHSPISNMVMNHNHSMTN 180
DB 194 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLKWTVPSPHSPISNMVMNHNHSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSS---LGSSSLGSSGLGGVAAANL 236
DB 254 TGVSMNTLHSLMLKGLAP-AAQAQVETAAENGVMAMSSLGSSSLGSSGLGGVAAANL 312
QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 6
Q7TWFS PRELIMINARY; PRT; 393 AA.
ID Q7TWFS
AC Q7TWFS5;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE PE family protein.
DE PPE60 OR MB3505.
GN Mycobacterium bovis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Eglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248346; CAD95692.1; --
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 72.7%; Score 949.5; DB 16; Length 393;
Best Local Similarity 79.7%; Pred. No. 7e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;

QY 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
DB 74 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLEQAAVEEASD 120
DB 134 IEANQAAVSQMWGQDAEAMGYAATAATATATALLPPEADPLITNPGGLLEQAVAVEAID 193
QY 121 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLKWTVPSPHSPISNMVMNHNHSMTN 180
DB 194 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLKWTVPSPHSPISNMVMNHNHSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSS---LGSSSLGSSGLGGVAAANL 236
DB 254 TGVSMNTLHSLMLKGLAP-AAQAQVETAAENGVMAMSSLGSSSLGSSGLGGVAAANL 312
QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 5
O06341 PRELIMINARY; PRT; 393 AA.
AC O06341;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTCY13E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A5007161; AAK47941.1; --
DR TIGR; MT3582; --
DR GO; GO:0016020; C-membrane; IEA.
DR GO; GO:0008233; F-peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE_I_1; 1.
SQ SEQUENCE 393 AA; 39688 MW; 86F0B67798955511 CRC64;

Query Match 77.2%; Score 1008.5; DB 16; Length 393;
Best Local Similarity 84.6%; Pred. No. 6.6e-61;
Matches 208; Conservative 14; Mismatches 19; Indels 5; Gaps 2;

QY 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
DB 74 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLEQAAVEEASD 120
DB 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEADPLITNPGGLLEQAVAVEAID 193
QY 121 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLKWTVPSPHSPISNMVMNHNHSMTN 180
DB 194 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLKWTVPSPHSPISNMVMNHNHSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSS---LGSSSLGSSGLGGVAAANL 236
DB 254 SGVSMNTLHSLMLKGFAP-AAQAQVETAAENGVMAMSSLGSSSLGSSGLGGVAAANL 312
QY 237 GRAASV 242
DB 313 GRAASV 318


```
Db 284 KIGGKPTGALAPLAEFALHTPILGSEGLGGSVSAGIGRAGLV 327

RESULT 9
O53950 PRELIMINARY; PRT; 423 AA.
AC O53950;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PPE family protein.
GN RV1801 OR MIV049.23 OR MT1850.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Garsdale J., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.D., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Orntoft L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL020201; CAA17722.1; -
DR EMBL; AE007044; AAK46122.1; -
DR PIR; B70931; B70931.
DR TIGR; MT1850; -.
DR TubercuList; RV1801; -.
DR InterPro; IPR00030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 423 AA; 41477 MW; 26E52CC271FBBF57 CRC64;

Query Match 33.8%; Score 441; DB 16; Length 423;
Best Local Similarity 41.2%; Pred. No. 3.5e-22;
Matches 117; Conservative 30; Mismatches 67; Indels 70; Gaps 9;

QY 1 VAMSVTAGQAEHTAAQVRVAAAYETAYGLTVPPPPVIAENRAELMILIAINLLGQNTPA 60
Db 72 VAMSVATAVHAELAGAQARLAIAAYEAFAATVPPVIAANRAQLMVLIAINFGQNTPA 131
QY 61 IAVNEAYGEMWAQDAAMFGYAAATATATATALLPFEAEPMISAGLLEQAAVEEASD 120
Db 132 IMWTEAQYEMWAQDAAMFGYAAATATATALLPFEAEPMISAGLLEQAAVEEASD 190
QY 121 TAAAN-----QLMNVVPAQLKOLAQPT----QGTFP-----SSKLGGLWKTVP 160
Db 191 TAAGGNLQSAFPQLLSAVPRALQGLALPTASQSASATPQWVDLGNLSTFLGG--AVTGP 248
QY 161 HRSPISNMVSMANNHMTNSGVMTNTLSMLKGFAPAAAQAVQTAQAQNVGRAMSSIG 220
Db 249 YTFP-----GVLPPSGVYLLGIQSVL-----VTQNGQGVSAALLG 283
QY 221 S-----SLGSSGLGGG--VAANLGRAASV 242
Db 284 KIGGKPTGALAPLAEFALHTPILGSEGLGGSVSAGIGRAGLV 327

RESULT 11
Q7TZJ3 PRELIMINARY; PRT; 393 AA.
AC Q7TZJ3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PPE family protein.
GN PPE26 OR MB1817.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
```

RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin J., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
RT "The complete genome sequence of *Mycobacterium bovis*."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547A8AB0 CRC64;

Query Match 33.1%; Score 432.5; DB 16; Length 393;
Best Local Similarity 41.0%; Pred. No. 1.2e-21;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;

QY 1 VAMSVTAQQAELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
Db 73 VAMSAAAAQAEQAATQARAAAAAFAAFAATVPPLIAANRASLMQLISTNVFGQNTSA 132
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPSEAPMTSAGLLEQAAVEASD 120
Db 133 IAAEAQYGEWMAQDAAMFYAGSSASASA-VTFPSTPPQIANPTAQTQAAAATAAG 191
QY 121 TAAA--NQLMNNVPOALKQAQQTGTPSSKLGGLWK-----TVSPH 161
Db 192 TAQSTLTEMITGLPNALQSLTSPQLQSS-NGPLSWLWQILFGTFNFTSISALLTDLPY 250
QY 162 RSPISNMVSMANNHSMVNTSGVMTNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLS 221
Db 251 ASFFYNTGELPYFSGMGNFQIAKTL-GLIGSAAPAAVA-----AAGDAKGLPLGLG 304
QY 222 SLGSSGLGGVAANLGRAASV 242
Db 305 MLG-----GGPVAAGLGNRAASV 321

RESULT 12
OS3939
ID OS3939 PRELIMINARY; PRT; 411 AA.
AC OS3939
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1789 OR MT1838 OR MTW049.11.
OS *Mycobacterium tuberculosis*.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brocks R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala E.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Davlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Skelton J.E., Taylor K., Whitehead S., Barrell B.G.,
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., L.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL022021; CAAL17711.1; ALT INIT.
DR EMBL; AE007043; AAK46108.1; -.
DR PIR; G70929; G70929.
DR TIGR; MT1838; -.
DR Tuberculist; Rv1789; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF0823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 33.1%; Score 432.5; DB 16; Length 411;
Best Local Similarity 41.0%; Pred. No. 1.3e-21;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;

QY 1 VAMSVTAQQAELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
Db 91 VAMSAAAAQAEQAATQARAAAAAFAAFAATVPPLIAANRASLMQLISTNVFGQNTSA 150
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPSEAPMTSAGLLEQAAVEASD 120
Db 151 IAAEAQYGEWMAQDAAMFYAGSSASASA-VTFPSTPPQIANPTAQTQAAAATAAG 209
QY 121 TAAA--NQLMNNVPOALKQAQQTGTPSSKLGGLWK-----TVSPH 161
Db 210 TAQSTLTEMITGLPNALQSLTSPQLQSS-NGPLSWLWQILFGTFNFTSISALLTDLPY 268
QY 162 RSPISNMVSMANNHSMVNTSGVMTNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLS 221
Db 269 ASFFYNTGELPYFSGMGNFQIAKTL-GLIGSAAPAAVA-----AAGDAKGLPLGLG 322
QY 222 SLGSSGLGGVAANLGRAASV 242
Db 323 MLG-----GGPVAAGLGNRAASV 339

RESULT 13
Q9Z5K0
ID Q9Z5K0 PRELIMINARY; PRT; 421 AA.
AC Q9Z5K0
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Putative PPE protein (PPE-family protein).
GN ML1182 OR MLCB1701.08C.
OS *Mycobacterium leprae*.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Mungall K., Basham D., Brown D., Garnier T., Churcher C., Harris D.,
RA Davies R.M., Davlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagels K., Macleod C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Skelton S., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
DR EMBL; AL049191; CAB39148.1; -.
DR EMBL; AL583921; CAC31563.1; -.
DR PIR; H87056; H87056.
DR Leptrama; ML1182; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF0823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 421 AA; 43119 MW; 680F5DFDCB65B4A7 CRC64;

Query Match 32.5%; Score 424.5; DB 16; Length 421;

Query Match	32.3%;	Score 422;	DB 16;	Length 409;	
Best Local Similarity	42.2%;	Pred. No. 6.6e-21;			
Matches 108;	Conservative 34;	Mismatches 92;	Indels 22;	Gaps 8;	
QY	1	VAMWVTAQAGBELTAAQURVAAAYETAYGLTVPVPVIAENRAELMILLIATNLLGQNTPA	60		
Db	73	VAMISVTAGQAEQAGAKIATGAYETAFATVPVPVIAENRALLMSLVAINIFGQNTPA	132		
QY	61	IAVNEAEYGEWMAQDAAMFYCAATATATATLPPFEAPEMTSGAGLLLQQAAYEASD	120		
Db	133	IAATEAHYAEMWADAAMFYAGSSATA-SQLAFFSEPTQTPSATAQSAVVAQAAQ	191		
QY	121	TAAA-----NQLMNNVPQALKQLAQPTQTPSSKGLGWKTIVSPHRSPINMYSMAN	173		
Db	192	AAASDDITAQLSQLISILPSTLQSLA--TTATATSAG--WDTV--LQSITILLANLT	244		
QY	174	NHMSWTNAGVS-----MTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSI--GSSLGS-S	226		
Db	245	GYSTIIGLGAIPGGWMLTGGQILGLAQNAPQVAVALLGFKAAAGNLSPLAPLRGGYIGDIT	304		
QY	227	GLGGGVAAANLGRAASV	242		

RESULT 15
Q7TZH7 PRELIMINARY; PRT: 409 AA.
ID Q7TZH7;
AC AC
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PPE family protein.
DE PPE32 OR M61837.
GN GN
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1765;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=AF2122797;
RC MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Dutchy S., Gironde S., Lacroix C., Monsempé C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RE ENBL; BX248340; CAD94540.1; -.
KW Complete proteome.
SQ SEQUENCE 409 AA; 39931 MW; D5789262B8B131A9E CRC64;
Query Match 32.2%; Score 420; DB 16; Length 409;
Best Local Similarity 41.6%; Pred. No. 9.1e-21;
Matches 107; Conservative 32; Mismatches 94; Indels 24; Gaps 8;
Qy 1 VAWMSVTAQAQLTAQVRVAAAAYETAYGLTVPPVPVIAENRAELMLIATNLGQNTTPA 60
Dd 73 VAWISVTAGAQSAQAQAKIAAGVYETAFAATVPPPVEANRALLMSLVATNIFGQNTTPA 132
Qy 61 IAVNEAEYGMWAODAAAMFGYAATAATATATLLPFEAPDWTSGAGLLEQAAAAVEASD 120
Dd 133 IAATEAHYEAEMWAQAAAAAMYAGSATA-SQLAPFSEPPQTNPSTATAQSVAUQAAG 191
Qy 121 TAAA-----NOLMNVNQALKQLAQPTQGTFSSKLGGLWKTVSPHRSPISNMVSMAN 173
Dd 192 AASSDIITLAQLSQLISLPSTLQSLA--TTATATSASAG--WDIV---LQSITTILANLT 244
Qy 174 NHWSMTNSGV---SWMTNTLSMKGFAPAQAQVOTAQNGVRAMSSLGSGLCS----- 225
Dd 245 GPYSIIIGLGAIPGGWWLITPGQIL-GLIAQNAPGVAALLGPKAAAGALSPLAPRGYTADI 303
Qy 226 SCIGGGGVAANIICPAASV 242

Wed Jun 23 16:34:26 2004

us-09-886-349a-12.rspt

Page 8

Db : |||| : ||| |
304 TELGGGATGGIARAIYV 320

Search completed: June 22, 2004, 17:22:50
Job time : 22.5043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 48.0027 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-14
Perfect score: 1949
Sequence: 1 MYDFGALPEINSARWYACP.....SGVLVPRPYWPHSPAAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1949	100.0	391	2	AAW32381 Mycobacte
2	1949	100.0	391	2	AAW32449 Mycobacte
3	1949	100.0	391	2	AAW64335 Mycobacte
4	1949	100.0	391	2	AAW81702 M. tuberc
5	1949	100.0	391	2	AAW04778 Mycobacte
6	1949	100.0	391	2	AAV38989 M. tuberc
7	1949	100.0	391	2	AAV39132 M. tuberc
8	1949	100.0	391	4	AAU01888 M. tuberc
9	1949	100.0	391	5	AAE29707 Mycobacte
10	1949	100.0	391	5	AAE17571 Mycobacte
11	1949	100.0	596	2	AAV32070 Mycobacte
12	1949	100.0	596	5	AAE29710 Mycobacte
13	1949	100.0	596	5	AAE17574 Mycobacte
14	1949	100.0	599	5	AAU74599 Antigenic
15	1949	100.0	600	2	AAV32068 Mycobacte
16	1949	100.0	600	5	AAU74597 Antigenic
17	1949	100.0	723	7	ADA26354 Mycobacte
18	1949	100.0	729	4	AAO22142 Ra12-H9-3
19	1949	100.0	729	5	AAE29709 Mycobacte
20	1949	100.0	729	5	AAE17573 Mycobacte
21	1949	100.0	729	7	ADA26374 Mycobacte
22	1949	100.0	744	4	AAU01902 M. tuberc
23	1949	100.0	813	7	ADA26367 Mycobacte
24	1949	100.0	815	4	AAU01904 M. tuberc
25	1949	100.0	825	7	ADA26366 Mycobacte

26 1949 100.0 875 7 ADA26365
27 1949 100.0 930 5 AAE29731 Mycobacte
28 1949 100.0 930 7 ADA26364 Mycobacte
29 1949 100.0 1010 7 ADA26356 Mycobacte
30 1949 100.0 1016 7 ADA26370 M. bovis
31 1949 100.0 1022 7 ADA26369 Mycobacte
32 1949 100.0 1154 7 ADA26368 Mycobacte
33 1945 99.8 788 4 AAU01903
34 1944 99.7 394 2 AAU04779
35 1944 99.7 729 5 AAE29708 Mycobacte
36 1944 99.7 729 5 AAE17572 Mycobacte
37 1939 99.5 729 5 ADA26373 Mycobacte
38 1931 99.1 729 2 AAY32059 Mycobacte
39 1902.5 97.6 726 5 AAU74588 Antigenic
40 1652.5 84.8 396 2 AAW64337 Mycobacte
41 1652.5 84.8 396 2 AAW81704 M. tuberc
42 1652.5 84.8 396 2 AAY38991 M. tuberc
43 1652.5 84.8 396 2 AAY39134 M. tuberc
44 1583 81.2 393 6 ABU36968 Protein e
45 1583 81.2 393 6 ABU34531 Protein e

ALIGNMENTS

RESULT 1
AAW32381
ID AAW32381 standard; protein; 391 AA.
XX
AC AAW32381;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-9FL.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-MAR-1996; 96US-00820280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX
(CORI-) CORIXA CORP.
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX
WPI; 1997-192904/17.
XX
N-PSDB; AAT91455.
XX
New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
useful for diagnosis of M. tuberculosis infection.

Example 3; Page 150-152; 190pp; English.

A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M. tuberculosis antigen, TbH-9FL. The immunogenic polypeptide can be used to diagnose M. tuberculosis infection by forming complexes with specific antibodies in the sample. Fragments of DNA encoding the immunogenic polypeptide can be used as diagnostic primers or probes and agents that bind to the antigen,

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSAEMVAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 DB 1 MVDGALPPEINSAEMVAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120
 DB 61 SSAGLMVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120

QY 121 LIATNLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 180
 DB 121 LIATNLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 180

QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSSLGSSG 300
 DB 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 360
 DB 301 LGGVVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 DB 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

RESULT 2

AAW32449
 ID AAW32449 standard; protein; 391 AA.

XX AAW32449;

XX 09-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen TbH-9FL.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX WO9709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US014674.

XX 01-SEP-1995; 95US-00523436.

XX 22-SEP-1995; 95US-00533634.

XX 22-MAR-1996; 96US-00620874.

XX 05-JUN-1996; 96US-00659683.

XX 12-JUL-1996; 96US-00680574.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;

XX Vedwick TH, Twardzik DR;

XX WPI; 1997-192903/17.

XX N-PSDB; AAT91521.

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.

XX Example 3; Page 138-139; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbH-9FL The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)

XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSAEMVAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60

DB 1 MVDGALPPEINSAEMVAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120

DB 61 SSAGLMVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120

QY 121 LIATNLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 180

DB 121 LIATNLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 180

QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240

DB 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSSLGSSG 300

DB 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 360

DB 301 LGGVVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

DB 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

RESULT 3

AAW64335
 ID AAW64335 standard; protein; 391 AA.

XX AAW64335;

XX 17-OCT-2003 (revised)

XX 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen TbH-9FL.

XX Tuberculosis; infection; diagnosis; antigen; TbH-9FL.

XX Mycobacterium tuberculosis; strain H37Rv.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018214.

XX 11-OCT-1996; 96US-00729622.

PR 13-MAR-1997; 97US-00818111.
XX (CORI-) CORIXA CORP.
FA
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ;
XX
XX WPI; 1998-251292/22.
DR N-PSDB; AAV44395.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
PT develop products for the detection of M. tuberculosis infection and
PT diagnosis of tuberculosis.
PS
PS Example 3; Page 133-135; 250pp; English.
XX
XX This polypeptide comprises Mycobacterium tuberculosis antigen TbH-9FL. It
CC is encoded by genomic DNA (see AAV44395) isolated from a M. tuberculosis
CC strain H37Rv genomic library using a probe from clone TbH-9 (see
CC AAV44371). The invention relates to compositions and methods for
CC diagnosing tuberculosis. It provides polypeptides (see AAV64291-W64379)
CC comprising an antigenic portion of a soluble M. tuberculosis antigen, or
CC an immunogenic portion of an M. tuberculosis antigen, as well as DNA
CC sequences encoding such polypeptides, recombinant expression vectors and
CC transformed or transfected host cells. Also claimed are methods and
CC diagnostic kits for detecting M. tuberculosis infection in a patient
CC using these polypeptides, antibodies or oligonucleotide probes and
CC primers for the diagnosis of tuberculosis. (Updated on 17-OCT-2003 to
CC standardise OS field)
XX
XX
SQ Sequence 391 AA;
Query Match 100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MVDGALPPEINARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMAHNMHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSG 300
DB 241 MVSMAHNMHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
RESULT 4
AAW81702
ID AAW81702 standard; protein; 391 AA.
XX
XX
AC AAW81702;
XX
XX 27-JAN-1999 (first entry)
DT
XX

M. tuberculosis immunogenic polypeptide TbH-9FL.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US018293.
XX
XX 11-OCT-1996; 96US-00730510.
PR 13-MAR-1997; 97US-00818112.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX Vedvick TS, Twardzik DR, Lodes MJ;
XX
XX WPI; 1998-261042/23.
DR N-PSDB; AAV64503.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
PT develop products for the detection of M. tuberculosis infection and for
PT diagnosis, treatment and prevention of tuberculosis.
XX
XX Example 3B; Page 128-129; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
CC inducing protective immunity against tuberculosis (TB). This sequence can
CC be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis
XX
SQ Sequence 391 AA;
Query Match 100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MVDGALPPEINARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMAHNMHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSG 300
DB 241 MVSMAHNMHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
RESULT 5

AAY04778
 ID AAY04778 standard; protein; 391 AA.
 AC AAY04778;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species protein sequence 5R.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 OS Mycobacterium sp.
 XX
 PN WO9909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR001813.
 XX
 PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX
 PA (INSP) INST PASTEUR.
 XX
 XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 PI
 DR WPI; 1999-181045/15.
 DR N-PSDB; AAX34030.
 XX
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 PS Claim 32; Fig 5R; 309pp; French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPPEINARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 DB 1 MVDFGALPPPEINARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 QY 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAG 180
 DB 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAG 180
 QY 181 LLEQAAVEASDTAAANQLMNNVPOALQOQAOPTQGTTPSSKLGGLWKTVPSPISN 240
 DB 181 LLEQAAVEASDTAAANQLMNNVPOALQOQAOPTQGTTPSSKLGGLWKTVPSPISN 240
 QY 241 MVSMAHNSMTNSGVSWTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSIGSSG 300
 DB 241 MVSMAHNSMTNSGVSWTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSIGSSG 300
 QY 301 LGGVAAANTLGRAASVGSLSVPQAAANQAVTPAABALPLTSLTSAERGPQMLGGLPV 360
 DB 301 LGGVAAANTLGRAASVGSLSVPQAAANQAVTPAABALPLTSLTSAERGPQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391
 DB 361 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391
 RESULT 6
 AAY38989
 ID AAY38989 standard; protein; 391 AA.
 XX
 AC AAY38989;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis recombinant antigen protein TBH-9FL.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003265.
 XX
 PR 18-FEB-1998; 98US-00024753.
 PR 05-MAY-1998; 98US-00072596.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedwick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AA219093.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis.
 PS Example 3; Page 168-169; 323pp; English.
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPPEINARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 DB 1 MVDFGALPPPEINARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 QY 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPMTSAG 180
 DB 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPMTSAG 180
 QY 181 LLEQAAVEASDTAAANQLMNNVPOALQOQAOPTQGTTPSSKLGGLWKTVPSPISN 240
 DB 181 LLEQAAVEASDTAAANQLMNNVPOALQOQAOPTQGTTPSSKLGGLWKTVPSPISN 240
 QY 241 MVSMAHNSMTNSGVSWTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSIGSSG 300

Db 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAARGPQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAARGPQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 7
AAV39132
ID AAV39132 standard; protein; 391 AA.
XX AC AAV39132;
XX 05-NOV-1999 (first entry)
DT M. tuberculosis antigen Tbh-9FL amino acid sequence.
DE Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX OS Mycobacterium tuberculosis.
XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US0003268.
XX PR 18-FEB-1998; 98US-00025197.
XX PR 05-MAY-1998; 98US-00072967.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YAM, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX WPI; 1999-527405/44.
XX DR N-PSDB; AAZ19305.
XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
XX tests and protective or therapeutic vaccines or compositions.
XX Example 3; Page 123-124; 299pp; English.
XX The present invention describes polypeptides comprising an immunogenic
XX part of a Mycobacterium tuberculosis antigen (Ag). Also described are
XX vaccines and fusion protein containing M. tuberculosis Ag's. M.
XX tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
XX polypeptides fragments, can be used in pharmaceutical compositions or
XX vaccines to generate a protective or therapeutic immune response to M.
XX tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
XX Ag can induce proliferation of, or cytokine secretion by, T, B or natural
XX killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
XX to AAZ19460 and AAY39083 to AAY39225 are used in the exemplification of
XX the present invention
XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0

QY 1 MVDFGALPPEINARMYAGPGASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
Db 1 MVDFGALPPEINARMYAGPGASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASPYVAMWSVTAGQAEELTAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI 120

Db 61 SSAGLMVAASPYVAMWSVTAGQAEELTAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLLGQNTPALAVNEAEYGENWAOAAAAMFCGAAAAATATATLIPFEAPEMTSAGG 180
Db 121 LIATNLLGQNTPALAVNEAEYGENWAOAAAAMFCGAAAAATATATLIPFEAPEMTSAGG 180
QY 181 LLEQAAAVEEASDTAAANQNLMMNVPOALQLOAQTOCTTSSKLGGLWKTVPSPHRSPLN 240
Db 181 LLEQAAAVEEASDTAAANQNLMMNVPOALQLOAQTOCTTSSKLGGLWKTVPSPHRSPLN 240
QY 241 MYSVANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSGSSG 300
Db 241 MYSVANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAARGPQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAARGPQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 8
AAU01888
ID AAU01888 standard; protein; 391 AA.
XX AC AAU01888;
XX 29-AUG-2001 (first entry)
DE M. tuberculosis antigen Tbh9 (Mtb39A).
XX Tbh9; Mtb39A; antigen; vaccine; tuberculosis; AIDS;
KW acquired immunodeficiency disease.
XX OS Mycobacterium tuberculosis.
XX PN WO200124820-A1.
XX PD 12-APR-2001.
XX PF 10-OCT-2000; 2000WO-US028095.
XX PR 07-OCT-1999; 99US-0158338P.
XX PR 07-OCT-1999; 99US-0158425P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
XX WPI; 2001-290576/30.
XX DR N-PSDB; AAS03779.
XX Vaccinating against Mycobacteria infections in mammals using fusion
XX proteins comprising combinations of heterologous antigens.
XX Example 2; Page 151-152; 168pp; English.
XX The sequence represents Mycobacterium tuberculosis Tbh9 (also known as
XX Mtb39A), an M. tuberculosis antigen. Compositions comprising at least 2
XX heterologous antigens, as a fusion protein, and vectors expressing the
XX fusion proteins are used as vaccines to prophylactically immunise mammals
XX (especially humans) against infection by Mycobacteria. The compositions
XX contain at least 2 heterologous antigens that increase the serological
XX sensitivity of individuals infected with tuberculosis, a disease
XX frequently affecting patients with acquired immunodeficiency disease,
XX AIDS
XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 4; Length 391;

Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MYDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 61 SSAGLWVAASPYVWMSVTAGQAELETAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLWVAASPYVWMSVTAGQAELETAQVRAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLPPFEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLPPFEAPEMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
DB 241 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAARERFGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAARERFGQMLGGLPV 360

QY 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
DB 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391

RESULT 9
AAE29707
ID AAE29707 standard; protein; 391 AA.
XX
AC AAE29707;
XX
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium sp. TbH9FL antigenic protein.
XX
KW vaccine; immunity; diagnostic agent; gene therapy; TbH9FL antigen.
XX
OS Mycobacterium sp.
XX
FN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
FF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WPI; 2002-759844/82.
XX
DR N-PSDB; AAD47082.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX tuberculosis.
XX
PS Disclosure; Page 85-87; 155pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a
XX fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX polynucleotide sequence encoding an antigen or an antigenic fragment from
XX Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX polypeptide or its fragment. The Leishmania polynucleotide is selected

CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC Mycobacterium sp. TbH9FL antigenic protein
XX
SQ Sequence 391 AA;
Query Match 100.0%; Score 1949; DB 5; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MYDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 61 SSAGLWVAASPYVWMSVTAGQAELETAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLWVAASPYVWMSVTAGQAELETAQVRAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLPPFEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLPPFEAPEMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
DB 241 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAARERFGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAARERFGQMLGGLPV 360

QY 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
DB 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391

RESULT 10
AAE17571
ID AAE17571 standard; protein; 391 AA.
XX
AC AAE17571;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species MTB39 (TbH9) protein #2.
XX
KW Fusion protein; antigen; serological sensitivity; immune response;
XX tuberculosis; infection; vaccine; MTB39; TBH9 protein.
XX
OS Mycobacterium sp.
XX
FN WO200198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US019959.
XX
PR 20-JUN-2000; 2000US-00597796.
XX
PR 01-FEB-2001; 2001US-0265737P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Alderson M;
XX
DR WPI; 2002-147798/19.

CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB59F (TbH9-Ra35) fusion protein
 XX
 SQ Sequence 596 AA;

Query Match 100.0%; Score 1949; DB 5; Length 596;
 Best Local Similarity 100.0%; Pred. No. 2.1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGIG 60
 DB 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGIG 68
 QY 61 SSAGLMVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 69 SSAGLMVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
 DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
 QY 181 LLEQAAVEEASDTAAANQLMNVPQALQQAQPTGTTTSSKLGLWKTVPSPHRSPISN 240
 DB 189 LLEQAAVEEASDTAAANQLMNVPQALQQAQPTGTTTSSKLGLWKTVPSPHRSPISN 248
 QY 241 MYSMANNHMTNSGVSMNTLTSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSG 300
 DB 249 MYSMANNHMTNSGVSMNTLTSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSG 308
 QY 301 LGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
 DB 309 LGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 DB 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 14
 AAU74599
 ID AAU74599 standard; protein; 599 AA.
 AC AAU74599;
 XX
 AC AAU74599;
 XX
 DT 29-AUG-2003 (revised)
 DT 08-MAY-2002 (first entry)
 DE
 DE Antigenic fusion protein Tb59-Ra35 (Mcb59f).
 XX
 XX Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; 1559-Ra35; Mcb59f.
 XX
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 XX Key Location/Qualifiers
 PH
 FT Misc-difference 597
 FT /label= OTHER
 FT /note= "OTHER= Xaa. Xaa= In frame stop codon"
 XX
 XX US2002009459-A1.
 XX
 XX 24-JAN-2002.

XX 07-APR-1999; 99US-00287849.
 PF
 XX 13-MAR-1997; 97US-00818112.
 PR 01-OCT-1997; 97US-00942578.
 PR 18-FEB-1998; 98US-00025197.
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.
 XX
 XX Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 XX WPI; 2002-171134/22.
 DR N-PSDB; ABK14139.
 DR
 XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 FT diagnosing, treating or preventing M. tuberculosis infection,
 FT particularly as vaccine for treating or preventing tuberculosis.
 XX
 PS Claim 1; Fig 12; 62pp; English.
 XX
 CC The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention. (Updated on 29-AUG-2003 to standardise OS field)
 XX
 SQ Sequence 599 AA;

Query Match 100.0%; Score 1949; DB 5; Length 599;
 Best Local Similarity 100.0%; Pred. No. 2.1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGIG 60
 DB 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGIG 68
 QY 61 SSAGLMVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 69 SSAGLMVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
 DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
 QY 181 LLEQAAVEEASDTAAANQLMNVPQALQQAQPTGTTTSSKLGLWKTVPSPHRSPISN 240
 DB 189 LLEQAAVEEASDTAAANQLMNVPQALQQAQPTGTTTSSKLGLWKTVPSPHRSPISN 248
 QY 241 MYSMANNHMTNSGVSMNTLTSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSG 300
 DB 249 MYSMANNHMTNSGVSMNTLTSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSG 308
 QY 301 LGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
 DB 309 LGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 DB 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 15

AAY32068
 ID AAY32068 standard; protein; 600 AA.
 XX AC AAY32068;
 XX DT 17-JAN-2000 (first entry)
 XX DE Mycobacterium tuberculosis antigen fusion protein Mtb61f.
 XX KW Tuberculosis; antigen; fusion protein; Mtb61f, TbH9; DPV; MTI; diagnosis;
 XX KW therapy; vaccine; immunogen.
 XX OS Mycobacterium tuberculosis.
 XX PN WO9951748-A2.
 XX PD 14-OCT-1999.
 XX PF 07-APR-1999; 99WO-US007717.
 XX PR 07-APR-1998; 98US-00056556.
 XX PR 30-DEC-1998; 98US-00223040.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skeiky YAW, Alderson M, Campos-Neto A;
 XX DR WPI; 1999-601610/51.
 XX DR N-PSDB; AAZ20203.
 XX PT New fusion proteins useful for diagnosis, prevention and treatment of
 XX PT tuberculosis.
 XX PS Claim 1; Fig 10A-B; 83pp; English.
 XX CC This sequence represents a recombinant Mycobacterium tuberculosis tri-
 CC antigen fusion protein, termed Mtb61f, composed of the antigens TbH9, Dpv
 CC and MTI. The fusion protein is expressed in host cells using a vector
 CC carrying a polynucleotide (see AAZ20203) comprising the coding sequences
 CC for the 3 antigens. The invention provides fusion proteins (see AAY32059-
 CC 71) containing at least 2 M. tuberculosis antigens. The new fusion
 CC proteins and polynucleotides encoding them are useful as vaccines for
 CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
 CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
 CC monitoring of disease progression, and treatment of tuberculosis. They
 CC are more effective immunogens than mixtures of the individual protein
 CC components
 XX SQ Sequence 600 AA;
 Query Match 100.0%; Score 1949; DB 2; Length 600;
 Best Local Similarity 100.0%; Pred. No. 2.1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDGALPPEINGARMYAGSGASLVAAQMDSDVASDLFSAASAFQSVVWGLTVGSMIG 60
 DB 9 MVDGALPPEINGARMYAGSGASLVAAQMDSDVASDLFSAASAFQSVVWGLTVGSMIG 68
 QY 61 SSAGLVAAASPVVAMWSTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
 DB 69 SSAGLVAAASPVVAMWSTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 180
 DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 188
 QY 181 LLEQAAVEEASDPAANQLMNNVPCALQOQTGGTTPSSKLGGLWKTVPSPHSPISN 240
 DB 189 LLEQAAVEEASDPAANQLMNNVPCALQOQTGGTTPSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSILGSSG 300
 DB 249 MYSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSILGSSG 308

QY 301 LGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAARERFGQMLGGLPV 360
 DB 309 LGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAARERFGQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
 DB 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399

Search completed: June 22, 2004, 17:15:29
 Job time : 49.0027 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 13.9576 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-14

Perfect score: 1949

Sequence: 1 MVDFGALPPEINSARWAGP.....SGVLRVPRPYWPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits: satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2.6/prodata2/iaa/5A_COMB.pep:*
- 2: /cgn2.6/prodata2/iaa/5B_COMB.pep:*
- 3: /cgn2.6/prodata2/iaa/6A_COMB.pep:*
- 4: /cgn2.6/prodata2/iaa/6B_COMB.pep:*
- 5: /cgn2.6/prodata2/iaa/PCTUS_COMB.pep:*
- 6: /cgn2.6/prodata2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	100.0	391	3	US-08-818-112-107
2	1949	100.0	391	4	US-08-818-111-102
3	1949	100.0	391	4	US-09-056-556-107
4	1949	100.0	391	4	US-09-072-596-102
5	1949	100.0	391	4	US-09-072-967-107
6	1949	100.0	596	4	US-09-287-849-26
7	1949	100.0	600	4	US-09-287-849-22
8	1944	99.7	729	4	US-09-223-040-2
9	1944	99.7	729	4	US-09-287-849-2
10	1552.5	84.8	396	3	US-08-818-112-111
11	1552.5	84.8	396	4	US-08-818-111-106
12	1552.5	84.8	396	4	US-09-056-556-111
13	1552.5	84.8	396	4	US-09-072-596-106
14	1552.5	84.8	396	4	US-09-072-967-111
15	1486.5	76.3	359	3	US-08-818-112-109
16	1486.5	76.3	359	4	US-08-818-111-104
17	1486.5	76.3	359	4	US-09-056-556-109
18	1486.5	76.3	359	4	US-09-072-596-104
19	1486.5	76.3	359	4	US-09-072-967-109
20	1187.5	60.9	358	4	US-09-287-849-8
21	1187	60.9	263	3	US-08-818-113-91
22	1187	60.9	263	4	US-08-818-111-92
23	1187	60.9	263	4	US-09-056-556-91
24	1187	60.9	263	4	US-09-072-596-92
25	1187	60.9	263	4	US-09-072-967-91
26	766.5	39.3	400	4	US-09-073-009-126
27	766.5	39.3	400	4	US-09-073-010-126

Sequence 16, Appl
Sequence 12, Appl
Sequence 142, App
Sequence 142, App
Sequence 131, App
Sequence 15, Appl
Sequence 15, Appl
Sequence 57, Appl
Sequence 208, App
Sequence 92, Appl
Sequence 114, App
Sequence 109, App
Sequence 114, App
Sequence 103, App
Sequence 114, App
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/POCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-107

Query Match 100.0%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLMVAASPPYVAMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLMVAASPPYVAMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPAIVANEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 121 LIATNLLGONTPAIVANEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 240
Db 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 240
QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 300
Db 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 300
QY 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTAAARALPLTSLTSAARERGPGOMLGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTAAARALPLTSLTSAARERGPGOMLGLPV 360
QY 361 GOMGARAGGGLSGVLVPPPPYVMPHSPAAG 391
Db 361 GOMGARAGGGLSGVLVPPPPYVMPHSPAAG 391

RESULT 2

US-08-818-111-102
; Sequence 102, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLMVAASPPYVAMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLMVAASPPYVAMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPAIVANEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 121 LIATNLLGONTPAIVANEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 240
Db 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 240
QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 300
Db 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 300
QY 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTAAARALPLTSLTSAARERGPGOMLGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTAAARALPLTSLTSAARERGPGOMLGLPV 360
QY 361 GOMGARAGGGLSGVLVPPPPYVMPHSPAAG 391
Db 361 GOMGARAGGGLSGVLVPPPPYVMPHSPAAG 391

RESULT 3

US-09-056-556-107
; Sequence 107, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-056-556-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAASAPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASAPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEKWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEKWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEASDTAAANLNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
DB 181 LLEQAAVEASDTAAANLNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
QY 241 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAVMSLSSLSGSSG 300
DB 241 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAVMSLSSLSGSSG 300
QY 301 LGGVAAANLGRAASVGSLSVFOAANAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 301 LGGVAAANLGRAASVGSLSVFOAANAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391

RESULT 4

US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAASAPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASAPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEKWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEKWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEASDTAAANLNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
DB 181 LLEQAAVEASDTAAANLNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
QY 241 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAVMSLSSLSGSSG 300
DB 241 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAVMSLSSLSGSSG 300
QY 301 LGGVAAANLGRAASVGSLSVFOAANAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 301 LGGVAAANLGRAASVGSLSVFOAANAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391

RESULT 5

US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967

[illegible]

QY	61	SSAGLMVAAASPYYAMNSVTTAGQAEILTAQQRVRAAAAYETAYGLTVPPPVIAENRAELMI	120
Db	202	SSAGLMVAAASPYYAMNSVTTAGQAEILTAQQRVRAAAAYETAYGLTVPPPVIAENRAELMI	261
QY	121	LIATNLLGONTPAITAVNEAYEGSWWQDAAAMFGYAAATATATATALLPPEEAPETVSAGG	180
Db	262	LIATNLLGONTPAITAVNEAYEGSWWQDAAAMFGYAAATATATATALLPPEEAPETVSAGG	321
QY	181	LLEQAAAVEEASDTAAANQLMNNVPQALQOAOPTCGTTTSSKLGGLMKTVSPHRSPISN	240
Db	322	LLEQAAAVEEASDTAAANQLMNNVPQALQOAOPTCGTTTSSKLGGLMKTVSPHRSPISN	381
QY	241	MVSMANNHMSMTNSGVSWTNTLSSMLKGFAPAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG	300
Db	382	MVSMANNHMSMTNSGVSWTNTLSSMLKGFAPAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG	441
QY	301	LGGGVAANLGRAASVGLSVPQAWAANAQAVTPAAPRALPLTSLTSAABRGPCQMLGGLPV	360
Db	442	LGGGVAANLGRAASVGLSVPQAWAANAQAVTPAAPRALPLTSLTSAABRGPCQMLGGLPV	501
QY	361	QCMGARAGGSLGVLRVPPRPYVMPHSPAAG	391
Db	502	QCMGARAGGSLGVLRVPPRPYVMPHSPAAG	532
RESULT 9			
US-09-287-849-2			
; Sequence 2, Application US/09287849			
; Patent No. 6627198			
; GENERAL INFORMATION:			
; APPLICANT: Reed, Steven G.			
; APPLICANT: Skeiky, Yasir A.W.			
; APPLICANT: Dillon, Davin C.			
; APPLICANT: Alderson, Mark			
; APPLICANT: Campos-Neto, Antonio			
; APPLICANT: Corixa Corporation			
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens			
; TITLE OF INVENTION: and Their Uses			
; FILE REFERENCE: 014058-009020US			
; CURRENT APPLICATION NUMBER: US/09/287,849			
; CURRENT FILING DATE: 1999-04-07			
; PRIOR APPLICATION NUMBER: US 08/918,112			
; PRIOR FILING DATE: 1997-03-13			
; PRIOR APPLICATION NUMBER: US 08/942,578			
; PRIOR FILING DATE: 1997-10-01			
; PRIOR APPLICATION NUMBER: US 09/025,197			
; PRIOR FILING DATE: 1998-02-18			
; PRIOR APPLICATION NUMBER: US 09/056,556			
; PRIOR FILING DATE: 1998-04-07			
; PRIOR APPLICATION NUMBER: US 09/223,040			
; PRIOR FILING DATE: 1998-12-30			
; NUMBER OF SEQ ID NOS: 46			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 2			
; LENGTH: 729			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion			
US-09-287-849-2			
Query Match 99.7%; Score 1944; DB 4; Length 729;			
Best Local Similarity 99.7%; Pred. No. 1.8e-153;			
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0			
QY	1	MVDFGALPPEINSRMYAGPGSASLVAAQWMDVSADLFSAAASAFQSVVWGLITVGSWIG	60
Db	142	MVDFGALPPEINSRMYAGPGSASLVAAQWMDVSADLFSAAASAFQSVVWGLITVGSWIG	201
QY	61	SSAGLMVAAASPYYAMNSVTTAGQAEILTAQQRVRAAAAYETAYGLTVPPPVIAENRAELMI	120
Db	202	SSAGLMVAAASPYYAMNSVTTAGQAEILTAQQRVRAAAAYETAYGLTVPPPVIAENRAELMI	261

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTFEEAPEMTSAGG 180
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTFEEAPEMTSAGG 321
QY 181 LLEQAAVEEASDTAAANQNMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 322 LLEQAAVEEASDTAAANQNMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 381
QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSSG 300
DB 382 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSSG 441
QY 301 LGGVAAANLGRAASVGSLSVQAWAANQVTPAARALPLTSLTSAERPGQMLGLPV 360
DB 442 LGGVAAANLGRAASVGSLSVQAWAANQVTPAARALPLTSLTSAERPGQMLGLPV 501
QY 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
DB 502 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 532

RESULT 10

US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-112-111

Query Match 84.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60

DB 1 VVDFGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLWAAASPYVAWMSVTAGQAEELTAAQVRAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 61 SSAGLWAAASPYVAWMSVTAGQAEELTAAQVRAAAAYETAYGLTVPPVPIAENRAELMI 120
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTFEEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTFEEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQNMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSSG 296
DB 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSSG 299
QY 297 GSGGLGGVAAANLGRAASVGSLSVQAWAANQVTPAARALPLTSLTSAERPGQMLG 356
DB 300 GSGGLGGVAAANLGRAASVGSLSVQAWAANQVTPAARALPLTSLTSAERPGQMLG 359
QY 357 GLPVGQMGARAG--GGLSGVLRVPPRPYVMPHSPAAG 391
DB 360 GLPLQLTNSGGGFGVSNALRMPRAVYVMPRVPAAAG 396

RESULT 11

US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-111-106


```
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-106

Query Match      84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 1 VVDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAAASPPYVAAWMSVTAGQAEHTAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLMVAAASPPYVAAWMSVTAGQAEHTAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
Db 121 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180

QY 121 LLEQAAVEEASDTPAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 121 LLEQAAVEEASDTPAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MVSMMNNHMTNSGVSMTNTLSMLKGFAPAAAQVQTAQAQNGVRAMS ----LGSSL 296
Db 241 IVSMLNNHVSMTNSGVSMASTLHSMKGFAP-AAAQAVETAQAQNGVQAMSSLGSLGSSL 299

QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERPGQMLG 356
Db 300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERPGQMLG 359

QY 357 GLPYGQMGARAG--GGLSGVLVRPPRYVMPHSPAAG 391
Db 360 GLPLGQLTNSGGGFGVSNALRMPRAYVMPRVPAAAG 396

RESULT 14
US-09-072-967-111
; Sequence 111, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMEUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-967-111

Query Match      84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 1 VVDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAAASPPYVAAWMSVTAGQAEHTAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLMVAAASPPYVAAWMSVTAGQAEHTAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
Db 121 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180

QY 121 LLEQAAVEEASDTPAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 121 LLEQAAVEEASDTPAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MVSMMNNHMTNSGVSMTNTLSMLKGFAPAAAQVQTAQAQNGVRAMS ----LGSSL 296
Db 241 IVSMLNNHVSMTNSGVSMASTLHSMKGFAP-AAAQAVETAQAQNGVQAMSSLGSLGSSL 299

QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERPGQMLG 356
Db 300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERPGQMLG 359

QY 357 GLPYGQMGARAG--GGLSGVLVRPPRYVMPHSPAAG 391
Db 360 GLPLGQLTNSGGGFGVSNALRMPRAYVMPRVPAAAG 396

RESULT 15
US-08-818-112-109
; Sequence 109, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```


COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-818-112-109

Query Match 76.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.9e-116;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

Qy	1	MVDFGALPPPEINARMYAGPGSASLVAAQWDSVASDLPESAASAFQSVVWGLTVGSGWIG	60
Db	1	MVDFGALPPPEINARMYAGPGSASLVAAQWDSVASDLPESAASAFQSVVWGLTVGSGWIG	60
Qy	61	SSAGLMVAAAAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI	120
Db	61	SSAGLMVAAAAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMT	120
Qy	121	LIATNLIGQNTPALAVNEAYGEMWQDAAMFGYAAATATATATATLLPPEEAPMTSAGG	180
Db	121	LIATNLIGQNTPALAVNEAYGEMWQDAAMFGYAAATATATATATLLPPEEAPMTSAGG	180
Qy	181	LLEQAAVEEASDTAAANQNMNVPOALQLOAQTGTTPTSSKLGGLWKTVPSPHSPISN	240
Db	181	LLEQAAVEEASDTAAANQNMNVPOALQLOAQTGTTPTSSKLGGLWKTVPSPHSPISN	240
Qy	241	MVSMANNHMSVNTSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVAMSS----LGSSSL	296
Db	241	VSSIANNHMSMGTGVSMTNTLHSMKGLAP-AAAQAVETAENGWAMSSLSGSLGSSSL	299
Qy	297	GSSGLGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQOMLG	356
Db	300	GSSGLGAGVAANLGRAASVGSLSVPPAWAANQAVTPAARALPLTSLTSAATAPGHMIG	359

Search completed: June 22, 2004, 17:27:32
Job time : 14.9576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 36.2142 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-14
Perfect score: 1949
Sequence: 1 MVDGALPEINSARMYAGP.....SGVLRVPPRPYVMPHSPAAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	100.0	391	12	US-09-886-349A-14
2	1949	100.0	391	14	US-10-193-002-102
3	1949	100.0	391	14	US-10-084-843-107
4	1949	100.0	391	14	US-10-098-732A-14
5	1949	100.0	596	9	US-09-287-849-26
6	1949	100.0	596	12	US-09-886-349A-20
7	1949	100.0	596	14	US-10-359-460-26
8	1949	100.0	596	14	US-10-098-732A-20
9	1949	100.0	600	9	US-09-287-849-22
10	1949	100.0	600	14	US-10-359-460-22
11	1949	100.0	723	15	US-10-369-983-2
12	1949	100.0	729	12	US-09-886-349A-18
13	1949	100.0	729	14	US-10-098-732A-18
14	1949	100.0	729	15	US-10-369-983-21
15	1949	100.0	729	15	US-10-369-983-22

16	1949	100.0	813	15	US-10-369-983-15	Sequence 15, Appl
17	1949	100.0	825	15	US-10-369-983-14	Sequence 14, Appl
18	1949	100.0	875	15	US-10-369-983-13	Sequence 13, Appl
19	1949	100.0	930	14	US-10-098-732A-65	Sequence 65, Appl
20	1949	100.0	930	15	US-10-369-983-12	Sequence 12, Appl
21	1949	100.0	1010	15	US-10-369-983-4	Sequence 4, Appl
22	1949	100.0	1016	15	US-10-369-983-18	Sequence 18, Appl
23	1949	100.0	1022	15	US-10-369-983-17	Sequence 17, Appl
24	1949	100.0	1154	15	US-10-369-983-16	Sequence 16, Appl
25	1944	99.7	729	9	US-09-287-849-2	Sequence 2, Appl
26	1944	99.7	729	12	US-09-886-349A-16	Sequence 16, Appl
27	1944	99.7	729	14	US-10-359-460-2	Sequence 2, Appl
28	1944	99.7	729	14	US-10-098-732A-16	Sequence 16, Appl
29	1944	99.7	729	15	US-10-359-459-2	Sequence 2, Appl
30	1934	99.2	391	12	US-09-872-186-8	Sequence 8, Appl
31	1652.5	84.8	396	14	US-10-193-002-106	Sequence 106, App
32	1652.5	84.8	396	14	US-10-084-843-111	Sequence 111, App
33	1583	81.2	393	12	US-10-282-122A-62455	Sequence 62455, A
34	1583	81.2	393	12	US-10-282-122A-64892	Sequence 64892, A
35	1486.5	76.3	359	14	US-10-193-002-104	Sequence 104, App
36	1486.5	76.3	359	14	US-10-084-843-109	Sequence 109, App
37	1187.5	60.9	358	9	US-09-287-849-8	Sequence 8, Appl
38	1187.5	60.9	358	14	US-10-359-460-8	Sequence 8, Appl
39	1187	60.9	263	12	US-09-886-349A-12	Sequence 12, Appl
40	1187	60.9	263	14	US-10-193-002-92	Sequence 92, Appl
41	1187	60.9	263	14	US-10-084-843-91	Sequence 91, Appl
42	1187	60.9	263	14	US-10-098-732A-12	Sequence 12, Appl
43	766.5	39.3	400	9	US-09-073-009-126	Sequence 126, App
44	766.5	39.3	400	9	US-09-793-306-126	Sequence 126, App
45	742.5	38.1	405	12	US-10-282-122A-62027	Sequence 62027, A

ALIGNMENTS

RESULT 1
US-09-886-349A-14
; Sequence 14, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886.349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 (TBH9FL)
US-09-886-349A-14

Query Match	100.0%	Score	1949	DB	12	Length	391
Best Local Similarity	100.0%	Pred. No.	1e-145				
Matches	391	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MVDGALPEINSARMYAGPGASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG	60				
Db	1	MVDGALPEINSARMYAGPGASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG	60				
Qy	61	SSAGLWMAAASPYVAMKSVTAQAEELTAQAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI	120				
Db	61	SSAGLWMAAASPYVAMKSVTAQAEELTAQAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI	120				

121 LIATNLLGONTPAIAVNEAEGEMWQADAAAFMGYAAATATATATATLPPREEPEMTSAGG 180
121 LIATNLLGONTPAIAVNEAEGEMWQADAAAFMGYAAATATATATLPPREEPEMTSAGG 180
181 LLEQAAVEEASDTAAANQLMNNVPAQALQQAQPTQCTTSSKGLGLWKTVPSPHRSPI 240
181 LLEQAAVEEASDTAAANQLMNNVPAQALQQAQPTQCTTSSKGLGLWKTVPSPHRSPI 240
241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 300
241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 300
301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGQMLGGLPV 360
301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGQMLGGLPV 360
361 GOMGARAGGGLSGVLVPPRPYPVPHSPAAG 391
361 GOMGARAGGGLSGVLVPPRPYPVPHSPAAG 391

RESULT 2

US-10-193-002-102

; Sequence 102, Application US/10193002

; Publication No. US20030135026A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, Davin C.

; Campos-Neto, Antonia

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

; TUBERCULOSIS

; NUMBER OF SEQUENCES: 350

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/193,002

; FILING DATE: 10-Jul-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,596

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; TELECOMMUNICATION INFORMATION: 210121.417C9

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 102:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 391 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 102:

; US-10-193-002-102

Query Match 100.0%; Score 1949; DB 14; Length 391;
Best Local Similarity 100.0%; Pred. No. 1e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYDFGALPPEINSAARMYAGPGSASIVAAAQWMDSVASDLFSAASAFOSVWGLTVGSGWIG 60
Db 1 MYDFGALPPEINSAARMYAGPGSASIVAAAQWMDSVASDLFSAASAFOSVWGLTVGSGWIG 60
QY 61 SSAGLWMAAASPYVAVMSVTAGQAEELTAQVRAVAAAAAYETAYGLTVPPVTAENRAELMI 120
Db 61 SSAGLWMAAASPYVAVMSVTAGQAEELTAQVRAVAAAAAYETAYGLTVPPVTAENRAELMI 120
QY 121 LIATNLLGONTPAIAVNEAEGEMWQADAAAFMGYAAATATATATLPPREEPEMTSAGG 180
Db 121 LIATNLLGONTPAIAVNEAEGEMWQADAAAFMGYAAATATATATLPPREEPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPAQALQQAQPTQCTTSSKGLGLWKTVPSPHRSPI 240
Db 181 LLEQAAVEEASDTAAANQLMNNVPAQALQQAQPTQCTTSSKGLGLWKTVPSPHRSPI 240
QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 300
Db 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGQMLGGLPV 360
QY 361 GOMGARAGGGLSGVLVPPRPYPVPHSPAAG 391
Db 361 GOMGARAGGGLSGVLVPPRPYPVPHSPAAG 391

RESULT 3

US-10-084-843-107

; Sequence 107, Application US/10084843

; Publication No. US20030143243A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, Davin C.

; Campos-Neto, Antonio

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 355

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/084,843

; FILING DATE: 25-Feb-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,967

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

```

; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 107:
US-10-084-843-107

Query Match      100.0%; Score 1949; DB 14; Length 391;
Best Local Similarity 100.0%; Pred. No. 1e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAAASPYYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPYYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLTPPEEAPEMTSAGG 180
DB 121 LIATNLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLTPPEEAPEMTSAGG 180
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 QMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
DB 361 QMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 5
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match      100.0%; Score 1949; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
```

Db 9 MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 61 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 69 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAAATATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAAATATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248
QY 241 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
Db 249 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 308
QY 301 LGGGVAANLGRAASVGLSVFPQAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
Db 309 LGGGVAANLGRAASVGLSVFPQAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399

RESULT 6

US-09-886-349a-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-886-349a-20

Query Match 100.0%; Score 1949; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 9 MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 61 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 69 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAAATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

Db 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248
QY 241 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
Db 249 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 308
QY 301 LGGGVAANLGRAASVGLSVFPQAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
Db 309 LGGGVAANLGRAASVGLSVFPQAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399

RESULT 7

US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-26

Query Match 100.0%; Score 1949; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 9 MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 61 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 69 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAAATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248

Qy	241	MYSMANNHSMWNTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVPMSSLGSSLGSSG	300
Db	249	MYSMANNHSMWNTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVPMSSLGSSLGSSG	308
Qy	301	LGGSVAANLGRAASVGSLSVPOAANAANQVTPAARALPTLSLTSAAERPGQMLGGLPV	360
Db	309	LGGSVAANLGRAASVGSLSVPOAANAANQVTPAARALPTLSLTSAAERPGQMLGGLPV	368
Qy	361	GQMGAAGGGLSGVLVPRPVPMPHSPAAG	391
Db	369	GQMGAAGGGLSGVLVPRPVPMPHSPAAG	399

```

RESULT 8
US-10-098-732A-20
    , Sequence 20, Application US/10098732A
    , Publication No. US20030175294A1
    , GENERAL INFORMATION:
    , APPLICANT: Skeiky, Yasir
    , APPLICANT: Brannon, Mark
    , APPLICANT: Gudarian, Jeffrey
    , APPLICANT: Corixa Corporation
    , TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
    , TITLE OF INVENTION: Leishmania Antigen
    , FILE REFERENCE: 014058-012010US
    , CURRENT APPLICATION NUMBER: US/10/098,732A
    , CURRENT FILING DATE: 2003-04-29
    , PRIOR APPLICATION NUMBER: US 60/275,837
    , PRIOR FILING DATE: 2001-03-13
    , NUMBER OF SEQ ID NOS: 80
    , SOFTWARE: PatentIn Ver. 2.1
    , SEQ ID NO 20
    , LENGTH: 596
    , TYPE: PRT
    , ORGANISM: Artificial Sequence
    , FEATURE:
    , OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
    , OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59P)

```

Query Match	100.0%	Score	1949	DB	14	Length	596		
Best Local Similarity	100.0%	Pred. No.	1.8e-145						
Matches	391	Conservative	0	Mismatches	0	Indels	0	Gaps	0
Qy	1	MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVMGLTVGSWIG	60						
Db	9	MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVMGLTVGSWIG	68						
Qy	61	SSAGLWVAASPPYVAMSVTQAGAELTAAQVRVAAAAYETAYGTUTPPPVIAENRAELMI	120						
Db	69	SSAGLWVAASPPYVAMSVTQAGAELTAAQVRVAAAAYETAYGTUTPPPVIAENRAELMI	128						
Qy	121	LIATNLIGONTPATAVNEABEYGEWMAODAAAMFGYAAATATATATLTFPEEAPEMTSAGG	180						
Db	129	LIATNLIGONTPATAVNEABEYGEWMAODAAAMFGYAAATATATLTFPEEAPEMTSAGG	188						
Qy	181	LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPTSN	240						
Db	189	LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPTSN	248						
Qy	241	MVSMANNHMSMTNSGVSWTNTLSSVLKGFAPAAAAAQAVQTTAAQNGVRAMSSLGSSGSG	300						
Db	249	MVSMANNHMSMTNSGVSWTNTLSSVLKGFAPAAAAAQAVQTTAAQNGVRAMSSLGSSGSG	308						
Qy	301	LGGGVAANLGRAAASVGSLSVQFAWAAANQAVTPAARALPLTSLTSAABRGQOMLGGPLVP	360						
Db	309	LGGGVAANLGRAAASVGSLSVQFAWAAANQAVTPAARALPLTSLTSAABRGQOMLGGPLVP	368						
Qy	361	QMGARAGGGLSGVLVRPPRPYVMPHSPAAQ	391						
Db	369	QMGARAGGGLSGVLVRPPRPYVMPHSPAAQ	399						

361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 3999

369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 3999

```

RESULT 9
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

```

Query Match	100.0%;	Score	1949;	DB	9;	Length	600;
Best Local Similarity	100.0%;	Pred. No.	1.18e-145;				
Matches	391;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
QY	1	MVDFGALPPPEINSARMYAGFSGASLVAAAQWMDVSADLPSAASAFQSVVMGLTVGSMIG	60				
Db	9	MVDFGALPPPEINSARMYAGFSGASLVAAAQWMDVSADLPSAASAFQSVVMGLTVGSMIG	68				
QY	61	SSAGLWVAASPYVAMSVTAGOAELETAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI	120				
Db	69	SSAGLWVAASPYVAMSVTAGOAELETAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI	128				
QY	121	LIATNLIGONTPAIVANEEYEGEMWAOADAAMFCGYAAATATATATLPPFEAPBMTSAGG	180				
Db	129	LIATNLIGONTPAIVANEEYEGEMWAOADAAMFCGYAAATATATATLPPFEAPBMTSAGG	188				
QY	181	LLEQAAVEEASDTAAANQLMNNVPQALQQLAQQTQGTTPSSKGLGLWKTVSPHRSPTS	240				
Db	189	LLEQAAVEEASDTAAANQLMNNVPQALQQLAQQTQGTTPSSKGLGLWKTVSPHRSPTS	248				
QY	241	MVSMANNHMSMTNSGVGKMTVLSNMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSSG	300				
Db	249	MVSMANNHMSMTNSGVGKMTVLSNMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSSG	308				
QY	301	LGGGVAANLGRASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARBGQMLGGLPV	360				
Db	309	LGGGVAANLGRASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARBGQMLGGLPV	368				
QY	361	GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG	391				
Db	369	GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG	399				

RESULT 10
US-10-359-460-22
; Sequence 22, Application US/10359460

```

; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/055,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; US-10-359-460-22

Query Match 100.0%; Score 1949; DB 14; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
Db 9 MVDGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 68
Qy 61 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 69 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
Qy 121 LIATNLLGQNTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATATLPPFEAPMTSAGG 180
Db 129 LIATNLLGQNTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATATLPPFEAPMTSAGG 188
Qy 181 LLEQAAVEEASDTAAANQNMNVPAQLQAQTOGTTPSSKLGGLWKTVPSPHSPIIN 240
Db 189 LLEQAAVEEASDTAAANQNMNVPAQLQAQTOGTTPSSKLGGLWKTVPSPHSPIIN 248
Qy 241 MVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVTAQAQNGVRAMSSSLGSSG 300
Db 249 MVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVTAQAQNGVRAMSSSLGSSG 308
Qy 301 LGGGVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGQMLGLPV 360
Db 309 LGGGVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGQMLGLPV 368
Qy 361 GQMGARAGGGLSGVLRVPPRPYPVPHSPAAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPYPVPHSPAAAG 399

RESULT 11
; US-10-369-983-2
; Sequence 2, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mutated
; OTHER INFORMATION: MYB32-MTB39F fusion protein (MTB32MutSA)
; US-10-369-983-2

Query Match 100.0%; Score 1949; DB 15; Length 723;
Best Local Similarity 100.0%; Pred. No. 2.3e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
Db 333 MVDGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 392
Qy 61 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 393 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 452
Qy 121 LIATNLLGQNTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
Db 453 LIATNLLGQNTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATLPPFEAPMTSAGG 512
Qy 181 LLEQAAVEEASDTAAANQNMNVPAQLQAQTOGTTPSSKLGGLWKTVPSPHSPIIN 240
Db 513 LLEQAAVEEASDTAAANQNMNVPAQLQAQTOGTTPSSKLGGLWKTVPSPHSPIIN 572
Qy 241 MVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVTAQAQNGVRAMSSSLGSSG 300
Db 573 MVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVTAQAQNGVRAMSSSLGSSG 632
Qy 301 LGGGVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGQMLGLPV 360
Db 633 LGGGVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGQMLGLPV 692
Qy 361 GQMGARAGGGLSGVLRVPPRPYPVPHSPAAAG 391
Db 693 GQMGARAGGGLSGVLRVPPRPYPVPHSPAAAG 723

RESULT 12
; US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1

```



```

; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
; OTHER INFORMATION: (Ral2-TbH9-Ra3MutSA)
US-09-886-349A-18

Query Match 100.0%; Score 1949; DB 12; Length 729;
Best Local Similarity 100.0%; Pred.No.2.3e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 MYDFGALPPEINSAHYAGPSASISLVAAQAQWDSVASDLFSAASAFQSVVWGLTYVCSWIG 60
DB 142 MYDFGALPPEINSAHYAGPSASISLVAAQAQWDSVASDLFSAASAFQSVVWGLTYVCSWIG 201
QY 61 SSAGLWMAAASPYYVAMSVTAQGAELTAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI 120
DB 202 SSAGLWMAAASPYYVAMSVTAQGAELTAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI 261
QY 121 LIATNLLGONTPAIIVNAEAYGEMWAQDAAMFGYAAATATATATATLIPPEEAPEMTSAG 180
DB 262 LIATNLLGONTPAIIVNAEAYGEMWAQDAAMFGYAAATATATATLIPPEEAPEMTSAG 321
QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGLWKTVPSPHRSPI 240
DB 322 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGLWKTVPSPHRSPI 381
QY 241 MYSMANNHSMWNTSGVSMNTLSSMLKGFAPAAAAQAVQTAAQNGVPMSSLSGSLGSSG 300
DB 382 MYSMANNHSMWNTSGVSMNTLSSMLKGFAPAAAAQAVQTAAQNGVPMSSLSGSLGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVPOWAAAANCAVTPAARALPLTSLTSAARPGQMLGGLPV 360
DB 442 LGGGVAANLGRAASVGSLSVPOWAAAANCAVTPAARALPLTSLTSAARPGQMLGGLPV 501
QY 361 GQMGARAGGLSGVLRVPPRPYPVMPHSPAAG 391
DB 502 GQMGARAGGLSGVLRVPPRPYPVMPHSPAAG 532

RESULT 13
US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
; OTHER INFORMATION: (Ral2-TbH9-Ra35MutSA)
US-10-098-732A-18

```

Qy	1	MVDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG	60
Db	142	MVDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG	201
Qy	61	SSAGLVAAASPYYVAMNSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVIAENRAELMI	120
Db	202	SSAGLVAAASPYYVAMNSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVIAENRAELMI	261
Qy	121	LIATNLLGQNTPAIVNAEAYEGEMWAQDAAMFGYAAAAATATATATLTPFEAPEMTSAGG	180
Db	262	LIATNLLGQNTPAIVNAEAYEGEMWAQDAAMFGYAAAAATATATLTPFEAPEMTSAGG	321
Qy	181	LLECAAAVEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTIVSPHRSPISN	240
Db	322	LLECAAAVEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTIVSPHRSPISN	381
Qy	241	MVSVANNHMSMTNSGYVSMNTLTSSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLGSSLGSSG	300
Db	382	MVSVANNHMSMTNSGYVSMNTLTSSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLGSSLGSSG	441
Qy	301	LGCGVAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAERGGQWLGLPLV	360
Db	442	LGCGVAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAERGGQWLGLPLV	501
Qy	361	GQMGARAGGGLSGVLRVPPRPYYMHPSPAAG	391
Db	502	GQMGARAGGGLSGVLRVPPRPYYMHPSPAAG	532
RESULT 14			
US-10-369-983-21			
; Sequence 21, Application US/10369983			
; Publication No. US20030235593A1			
; GENERAL INFORMATION:			
; APPLICANT: Skeiky, Yasir			
; APPLICANT: Guderian, Jeff			
; APPLICANT: Reed, Steven			
; APPLICANT: Corixa Corporation			
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis			
; FILE REFERENCE: 014058-009081US			
; CURRENT APPLICATION NUMBER: US/10/369,983			
; CURRENT FILING DATE: 2003-02-18			
; PRIOR APPLICATION NUMBER: US 60/357,351			
; PRIOR FILING DATE: 2002-02-15			
; NUMBER OF SEQ ID NOS: 22			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 21			
; LENGTH: 729			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F			
US-10-369-983-21			
Query Match 100.0%; Score 1949; DB 15; Length 729;			
Best Local Similarity 100.0%; Pred. No. 2,3e-145; Indels 0; Gaps 0;			
Matches 391; Conservative 0; Mismatches 0;			
Qy	1	MVDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG	60
Db	142	MVDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG	201
Qy	61	SSAGLVAAASPYYVAMNSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVIAENRAELMI	120
Db	202	SSAGLVAAASPYYVAMNSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVIAENRAELMI	261
Qy	121	LIATNLLGQNTPAIVNAEAYEGEMWAQDAAMFGYAAAAATATATLTPFEAPEMTSAGG	180
Db	262	LIATNLLGQNTPAIVNAEAYEGEMWAQDAAMFGYAAAAATATATLTPFEAPEMTSAGG	321
Qy	181	LLECAAAVEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTIVSPHRSPISN	240
Db	322	LLECAAAVEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTIVSPHRSPISN	381

Search completed: June 22, 2004, 18:07:55
Job time : 37.2142 secs

Qy	241	MYSMANHHMSMTNSGV	STNTLS	SMWKGFAPAAAAQAV	TAQNGVVRMS	SLGSS	LGSSG	300
Db	382	MYSMANHHMSMTNSGV	STNTLS	SMWKGFAPAAAAQAV	TAQNGVVRMS	SLGSS	LGSSG	441
Qy	301	LGGSVAANI	GRAASVGS	LSVPAQMAAANQAV	TPAARALP	ITS	SAABRGPGOMLGGLPV	360
Db	442	LGGSVAANI	GRAASVGS	LSVPAQMAAANQAV	TPAARALP	ITS	SAABRGPGOMLGGLPV	501
Qy	361	GQMGARAGGGLSGVLR	VPPRP	VPYMPHP	PAAG	391		
Db	502	GQMGARAGGGLSGVLR	VPPRP	VPYMPHP	PAAG	532		

```

RESULT 15
US-10-369-983-22
; Sequence 22, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983.
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; -OTHER INFORMATION: MTB72FmutSA (MTb72F-mutSA)
US-10-369-983-22

```

Query Match	100.0%;	Score	1949;	DB	15;	Length	729;
Best Local Similarity	100.0%;	Pred. No.	2.3e-145;				
Matches	391;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MVDFGALPPEINSARMYAGPGSASLVAQAQWDSVASDLFSAASAFQSVVMGLTVGSWIG	60				
Db	142	MVDFGALPPEINSARMYAGPGSASLVAQAQWDSVASDLFSAASAFQSVVMGLTVGSWIG	201				
QY	61	SSAGLMWAAASPPYVAMSVTAGQAEUATAAQVRVAAAAYETAYGLTTPPPVIAENRAELMI	120				
Db	202	SSAGLMWAAASPPYVAMSVTAGQAEUATAAQVRVAAAAYETAYGLTTPPPVIAENRAELMI	261				
QY	121	LIATNLGQNTPATAVNEASVGEWMAQDAAAMFGYAAATATATATATLLEFEEAPEMTSAGG	180				
Db	262	LIATNLGQNTPATAVNEASVGEWMAQDAAAMFGYAAATATATATLLEFEEAPEMTSAGG	321				
QY	181	LLEQAAAVEEASDTAAANQALMNNVPQALQQLAQCTQGTTTPSSKLGGLWKTVSPHRSIPSN	240				
Db	322	LLEQAAAVEEASDTAAANQALMNNVPQALQQLAQCTQGTTTPSSKLGGLWKTVSPHRSIPSN	381				
QY	241	MVSMANNHMTNSGVSGVMTNTLSMMLKGFAPAAAAAQAVQTAAQNGVRAMSSLGSSLGSSG	300				
Db	382	MVSMANNHMTNSGVSGVMTNTLSMMLKGFAPAAAAAQAVQTAAQNGVRAMSSLGSSLGSSG	441				
QY	301	LGGGVAANLGRASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGQCMGLGPV	360				
Db	442	LGGGVAANLGRASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGQCMGLGPV	501				
QY	361	QMGARAGGGLSGVLRVPPPPYVMPHSPAAG	391				
Db	502	QMGARAGGGLSGVLRVPPPPYVMPHSPAAG	532				

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 10.9397 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-14
Perfect score: 1949
Sequence: 1 MVDFGALPPEINSARMYAGP.....SCVLVRPPRYVMPHSPAAG 391
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1949	100.0	391	2	B70608 probable PPE prote
2	1656.5	85.0	396	2	H70741 probable PPE prote
3	1583	81.2	393	2	C70568 probable PPE prote
4	779.5	40.0	393	2	G70929 probable PPE prote
5	751	38.5	409	2	A70932 probable PPE prote
6	737.5	37.8	403	2	H70931 probable PPE prote
7	736	37.8	423	2	B70931 PPE-family protein
8	731	37.5	421	2	H87056 probable PPE prote
9	705	36.2	408	2	G70925 probable PPE prote
10	702	36.0	391	2	B70625 probable PPE prote
11	702	36.0	413	2	F70560 probable PPE prote
12	688.5	35.3	463	2	C70931 probable PPE prote
13	682.5	35.0	468	2	B70932 probable PPE prote
14	675	34.6	380	2	A70646 probable PPE prote
15	668.5	34.3	394	2	G70881 probable PPE prote
16	667.5	34.2	385	2	H70903 probable PPE prote
17	635	32.6	350	2	F70929 probable PPE prote
18	633.5	32.5	365	2	E70929 probable PPE prote
19	615.5	31.6	402	2	A70882 probable PPE prote
20	603	30.9	423	2	C70582 probable PPE prote
21	597.5	30.7	391	2	D70922 probable PPE prote
22	589	30.2	406	2	E70675 probable PPE prote
23	588	30.2	394	2	A70504 probable PPE prote
24	575.5	29.5	391	2	A70663 probable PPE prote
25	495.5	25.4	3300	2	D70575 probable PPE prote
26	473	24.3	180	2	G70834 probable PPE prote
27	452.5	23.2	3716	2	E70969 probable PPE prote
28	448.5	23.0	580	2	G70570 probable PPE prote
29	444	22.8	346	2	H70874 probable PPE prote

30	443	22.7	963	2	B70524 probable PPE prote
31	441	22.6	487	2	C70830 probable PPE prote
32	439.5	22.6	655	2	A70931 probable PPE prote
33	439.5	22.0	479	2	D70676 probable PPE prote
34	438.5	22.0	1053	2	B70987 probable PPE prote
35	436.5	21.9	443	2	C70780 probable PPE prote
36	433.5	21.7	2523	2	F70846 probable PPE prote
37	422.5	21.7	615	2	E70663 probable PPE prote
38	418	21.4	678	2	A70762 probable PPE prote
39	418	21.4	3157	2	B70969 probable PPE prote
40	415.5	21.3	618	2	F70552 probable PPE prote
41	415	21.3	582	2	H70675 probable PPE prote
42	414.5	21.3	987	2	E70808 probable PPE prote
43	412	21.1	590	2	F70946 probable PPE prote
44	412	21.1	645	2	F70825 probable PPE prote
45	409.5	21.0	409	2	A70647 probable PPE prote

ALIGNMENTS

RESULT 1

B70608

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C.Species: Mycobacterium tuberculosis

C.Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C.Accession: B70608

R.Cole, S.I.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A.Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whithead, S.; Barrell, B.G.

A.Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A.Reference number: A70500; MUID:98295987; PMID:9634230

A.Accession: B70608

A.Status: preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-391 <COL>

A.Cross-references: GB:Z93777; GB:AL123456; NID:g3261726; PIDN:CAB07839.1; PID:e311073; I

A.Experimental source: strain H37RV

C.Genetics:

A.Gene: PPE

Query Match 100.0%; Score 1949; DB 2; Length 391;

Best Local Similarity 100.0%; Pred. No. 3.5e-108;

Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASIVAAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60

Db 1 MVDFGALPPEINSARMYAGPGSASIVAAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60

QY 61 SSAGLWVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120

Db 61 SSAGLWVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120

QY 121 LIATNLLQNTPTAVNEAEYGEWQAQAAAMFGVAAATATATATLLPFEAPENTSAGG 180

Db 121 LIATNLLQNTPTAVNEAEYGEWQAQAAAMFGVAAATATATATLLPFEAPENTSAGG 180

QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSKLGLWKTVPSPHRSPISN 240

Db 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSKLGLWKTVPSPHRSPISN 240

QY 241 MYSMANNHMTNSGVSMNTLSNMLKGFAPAAAQAVQTAQNGVRAMSSLSGSLGSSG 300

Db 241 MYSMANNHMTNSGVSMNTLSNMLKGFAPAAAQAVQTAQNGVRAMSSLSGSLGSSG 300

QY 301 LGGGVAANLGRAASVGSLSVQAAANQAVTPAARALPLTSLTSAARFGQMLGLPLV 360

Db 301 LGGGVAANLGRAASVGSLSVQAAANQAVTPAARALPLTSLTSAARFGQMLGLPLV 360

QY 361 GOMGARAGGSLSGVLRVPPRYVMPHSPAAG 391

Db 361 GOMGARAGGSLSGVLRVPPRYVMPHSPAAG 391

Db 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
RESULT 2
H70741
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70741
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70741
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-396 <COL>
A:Cross-references: GB:275555; GB:AL123456; NID:g3261608; PIDN:CAA99966.1; PID:e250360;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE
Query Match 85.0%; Score 1656.5; DB 2; Length 396;
Best Local Similarity 85.1%; Pred. No. 6.7e-91;
Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;
Qy 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
Db 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
Qy 61 SSAGLWVAASPPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWVAASPPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Qy 121 LIATNLLGQNTPAIVNAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPEMTSAGG 180
Db 121 LIATNLLGQNTPAIVNAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPEMTSAGG 180
Qy 181 LLEQAAVEEASDTAAANQLMNNVPAQQLAQAOPTGTTSSKLGGLWKTVPSPHSPI 240
Db 181 LLEQAAVEEASDTAAANQLMNNVPAQQLAQAOPTGTTSSKLGGLWKTVPSPHSPI 240
Qy 241 MVSMANNHMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVRAMSS----LGSSL 296
Db 241 MVSMANNHMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVRAMSS----LGSSL 296
Qy 297 GSSGLGGVAAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERPGQMLG 356
Db 297 GSSGLGGVAAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERPGQMLG 356
Qy 357 GLPVQMGARAG--GGLSGVLRVPPRYVMPHSPAAG 391
Db 357 GLPVQMGARAG--GGLSGVLRVPPRYVMPHSPAAG 391
Qy 360 GLPLGH-SVNAAGSINNLRVPAAYALPTPTAAG 393
Db 360 GLPLGH-SVNAAGSINNLRVPAAYALPTPTAAG 393
RESULT 4
G70929
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17711.1; PID:e125460C
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE
Query Match 40.0%; Score 779.5; DB 2; Length 393;
Best Local Similarity 43.7%; Pred. No. 4.3e-39;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;
Qy 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 61
Db 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 61
Qy 62 SAGLWVAASPPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 121
Db 62 SAGLWVAASPPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 121
Qy 61 ASAAAEAVAPYVAVMSAAAQAEQATQARAAAAAFAAATVPPPLTAANRASLMQL 120
Db 61 ASAAAEAVAPYVAVMSAAAQAEQATQARAAAAAFAAATVPPPLTAANRASLMQL 120

QY 122 IATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAEMTISAGGL 181
Db 121 ISTNVFGONTSAIAAEAEYGEWMAQDSAAVAYAGSSASASA-VTFPSTPQIANPTAQ 179
QY 182 LEQAAAVEEASDTAAA--NOLMNNVFOALQOLAQPTGCTTPSSKLGGLWK----- 229
Db 180 GTQAAAATAAGTAQSTLTEMITGLPNALQSTSPILQSS--NGPLSWLWQILFGTNPFT 238
QY 230 -----TVSPHSPISNNMVMANNMSTNSGVSMNTLSSMLKGFAPAAAQAQVATAA 282
Db 239 SISALLTDLCQPYASFFYNTEGLPYFISGNGNFIQSAKTL-GLIGSAAFAVA-----AA 292
QY 283 QNGVRAMSSLGSSIGSGGLGGVAAVNLGRASVGSLSVQAAWAA--ANQAVTPAARALPLT 341
Db 293 GDAAKGLPGLGMLG---GGPVAAGLIGNAASVGLSVPPVMSGPLPGSVTPCAAFLPVS 348
QY 342 SLTSAAEERGPGOMLGLPVQCMGARAGGGLSVLVRPPPYVMPHPSPAAG 391
Db 349 TVSAAPAEAPGSLGGLPL-----AGAGGAGAGP-RYGFPTVMARPPFAG 393

RESULT 5
A70932
probable pPE protein- Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70932
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-409 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17729.1; PID:e125454
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 38.5%; Score 751; DB 2; Length 409;
Best Local Similarity 44.7%; Pred. No. 2.2e-37;
Matches 177; Conservative 53; Mismatches 138; Indels 28; Gaps 10;

QY 2 VDFGALPPEINARMYAGPGSASLVAAAQWDSVADLPSAASAFQSVWGLTVGWSIGS 61
Db 1 MDFGALPPEINSGRMYAGPGSGPLLAATAAAWDLAAELYSAGASYGSTIEGLTVAPWMPG 60
QY 62 SAGLWMAAASPYYAWMSVTAGAELTAAQVRAAAAAYETAYGLTVPPVPVIAENRAELMIL 121
Db 61 SSITMAAAVAPYVANI SVTAQAEQAGCAQAKIAGVYETAFATVPPVPIEANKALLMSL 120
QY 122 IATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAEMTISAGGL 181
Db 121 VATNIFGONTPTAIAATEAHYAEMWMAQDAAMFGYAGSSASA-SQLAPFSEPPPTNPST 179
QY 182 LEQAAAVEEASDTAAA-----NOLMNNVFOALQOLAQPTGCTTPSSKLGGLWKTVSPH 234
Db 180 AAQSAVVAQAGAAASDITQLSQLSLPSTLQSLA--TTATATSAAG--WDIV--- 232
QY 235 RSPISNNMVMANNMSTNSGVSMNTLSSMLKGFAPAAAQAQVATAAONGVRAM 289
Db 233 LQSITILLANTGPYSIIGLCAIPGGWMLTFGQILGLAQNAPGVAALLGPKAAAGALSPL 292
QY 290 SSL--GSSLGS--SGLGGVAAVNLGRASVGSLSVQAAWAA--ANQAVTPAARALPLTSLTSA- 346
Db 293 APLSGGVIGDITPLGGGATGTGATAIYVGSLSVPPVMSGPLPGSVTPCAAFLPVS 352
QY 347 AERPGQWMLGCLPVQCMGARAGGGL-----SGVLRV 377

Db 353 AAEAPGALFGEMALSSLAGRALAGTAVESGAGAAKV 388

RESULT 6

H70931

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70931

R:Cole, S.T.; Davies, R.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Hollroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: H70931

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-403 <COL>

A:Cross-references: GB:AL020221; GB:AL123456; NID:G3250699; PIDN:CAAL7728.1; PID:el25461f

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 37.8%; Score 737.5; DB 2; Length 403;

Best Local Similarity 42.6%; Pred No. 1.4e-36;

Matches 181; Conservative 60; Mismatches 123; Indels 61; Gaps 12;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQMDVSDLPSAASAFOSVWGLTVGSWIGS 61

DB 5 LDFATLPEINSARMYSGAGSAPMLAAASAMHGLSAELFASALSYSVSLTGTGEWHGP 64

QY 62 SAGLMVAASPYVAMWSVTAGAEHTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMIL 121

DB 65 ASASMTAAAPYVAMWSVTAVRERQAGAEAAAAYEAAPATVPPPVIEANRAQIMAL 124

QY 122 IATNLLGONTFAIVNEABYEGEMWQAQDAAMFGYAAATATATATLTPFEAPMTSAGGL 181

DB 125 IATNVLGONAPALAAATEACYAEWSQDAMMYGACASAAAT-QLTPFTPEPVQTINASGL 183

QY 182 LEQAAAVEASDTAA-----NOLMNVVPAQLQLOAQPTQGT-----TPSS 222

DB 184 AAQSAIAIAHATGASAGAQTTLQSILTAAPVSLQGLSSSTATFASGPSGLLGIIVSGSS 243

QY 223 KLGGIWKTVSPHRSPISNMVMANNHMTNSGVSM-TNTLS---SMLKGFAPAAAAQAV 278

DB 244 WLDKLWALLDPN-----SNFNWNTIASGGLFLPSNTIAPFLGLGGVAAAADAAGDV 293

QY 279 QTAANGVRAMSSLGSL-----GSGLCGGVAAANLGRAASVGSLSVPQMAAANQAVTP 333

DB 294 LGEATSG-----GLGGALVAPLGSAGGLGTVAAGLGNAAATVGLTSPSPSWTAAAPLASP 348

QY 334 AARAL---PLTSLTSAAEERGPGQMLGGLPVGOMGARAGGSLGVLRVPP---RPPVWPH 386

DB 349 LGSALGGTFWVAPPFAVAG-----MPGMFFGTMGGGGFG-----RAVPQYGFEPNFVAR 398

QY 387 SPAAG 391

DB 399 PPAAG 403

RESULT 7

H70931

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70931

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Hollroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

Db	121	VATNFFGQNTPAIAAEQAQAEWQAQAAAYAYAGSRAIAT-ELTPFAAPVTSPAAL	179
QY	182	LEQAAA-----VEEASDTAAANQLMNN-----VP--QALQQ-LAQPTGCTTPSSKL---	224

Db 180 AGQAAATVSTVPLATTAAPVQLQLSSTSLIPWYSLQWLAEMLLGLTPDNRMTIV 239
QY 225 -----GGLWKTVPSPHSPISSNVSMANNHMTNSGVSMNTLSSMLKGFAPAAA 275
Db 240 RLIGISYFDEGL-----LQFASLAQAQIPOTPGAG--DSGSSVLDSWGTIFA 287
QY 276 QAVCTAAQNGVRAMSL--GSSLSG-----SGLGGVAAANLGRAASVGSLS 319
Db 288 -----GPRASPVSAGGAVGGVQTPQYPMWALDRSIGSVSAALGKGSAGLS 338
QY 320 VPCWAAANQAVTPAARALP--LTSLSAAERPGQMLGGLPVGQMGARAGGSLGVLR 376
Db 339 VPPDMAARAWNPAAWRIPGDDVTALRGTAENA---LLRGPFMASAGOSTGGF--VHK 393
QY 377 VPPRPVYMPHSPAAG 391
Db 394 YGFLAVMQRPFPAG 408

RESULT 10
B70625
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70625
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70625
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-391 <COL>
A: Cross-references: GB:292539; GB:AL123456; NID:G3261714; PIDN: CAB06873.1; PID: e304546;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 391;
Best Local Similarity 42.3%; Pred. No. 1.6e-34;
Matches 166; Conservative 62; Mismatches 150; Indels 14; Gaps 7;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFOSVVMGLTVGSWIGS 61
Db 1 MDFGALPPEINSARMYAGAGAGPMMAAGAAWNGLAALAEELGTAAASYESVITRLTTESWMP 60
QY 62 SAGLMVAAASPYVAVWSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELML 121
Db 61 ASWAVAAQPYLANITYTAAAHAGSQMASAAAYEAAYAVTVPPEVVAANRALLAAL 120
QY 122 IATNLGONTPTAIVNEAYEGEYVMAQDAAMFGYAAATATATATLPPFEAPEMTSAGL 181
Db 121 VATNLGINTPTAMATEALYAEYVMAQDALAMYAAASG--AAGMLQLPSPPQTTNPGL 179
QY 182 LEQAAVEASDTPAANQ-----LMNVPCALQQLAQPTQGTTPSSKLGKLVKTVSPHRS 236
Db 180 AQSAVGAATAAANVQNSVADLSSLPNAVSGLSAPVTSVLDSTGSGIITADIALIA 239
QY 237 P--ISNVSMANNHMS--MTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSG 293
Db 240 TPVANIINSVNTAAWYNAIPTAIFLANALNSGAPVAIEGALEAAEG--AASAAA 296
QY 294 SSLGSSGLGGVAAANI GRAASVGSLSVPOQWAAANQAVTPAARALPLTSLTSAARPGQ 353
Db 297 AGLADSVTPAGLGSIGATLVGLRSLVPAWSTAAAPATTAGATALEGSGWTVAEEA-QP 355
QY 354 MCGGLPVQMGARAGGSLGVLRVPPRPVMP 385
Db 356 VTGWMP-GWASAAKGTGAVAGPRYGRKPTMP 386

RESULT 11
F70560
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70560
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70560
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-413 <COL>
A: Cross-references: GB:295436; GB:AL123456; NID:G3261770; PIDN: CAB08826.1; PID: e316585;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 413;
Best Local Similarity 39.1%; Pred. No. 1.7e-34;
Matches 168; Conservative 61; Mismatches 145; Indels 56; Gaps 7;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFOSVVMGLTVGSWIG 60
Db 1 MLDFAQLPPEVNSALMYAGPGSGPMLAAAWEALAAELQTTASTYDALITGLADGPWQ 60
QY 61 SSAGLMVAAASPYVAVWSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
Db 61 SSAASMVAAATPQVAVLNRSTAGQAEQAGSQAVAAASAYEAAFFATVPPPEIAANRALLMA 120
QY 121 LIATNLGONTPTAIVNEAYEGEYVMAQDAAMFGVAAATATATLPPFEAPEMTSAGG 180
Db 121 LIATNLGONTPTAIAATAEAQVAAEYVMAQDAAMFGVAAATATATLPPFEAPEMTSAGG 179
QY 181 LLEQAAVEASDTPAANQVLMNVPCALQQLAQPTQGTTPSSKLGKLVKTVSPHRSPTN 240
Db 180 LASQAAASVQAVSGAANAQALTDIPKAL-----FGLSGIFTNPPMLTDLGK 226
QY 241 MVSMAHHMTNSGVSMNTLSSMLKGFAPAA---AAQAVQT----- 280
Db 227 ALGTLGHTWSSDGGSLIVGGVGLDFVQGVTSAGSLDASVAMDTFGKVVSPARLMVTQPKD 286
QY 281 -----AAQNGVRAMSSLSGSSLSGSGVAAANLGRAASVGSLSVPOA 323
Db 287 YFGLAHLDPKWASEGAKAAGEAAKALPAAVPAIPISAGL-SCVAGAVGQAASVGLKVPVAV 345
QY 324 WAAANQAVTPAARALPLTSLTSAARPGQMLGSLPVQMGARAGGSLGVLR--RVPPRP 381
Db 346 WTATTPAASPVAALAAASNLGAAAAAEGSTHAFGGMPL--MGSGAGAFNNFAAPRYGPKP 403
QY 382 YVPHSPSPAAG 391
Db 404 TVIAQPPAGG 413

RESULT 12
C70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: C70932

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-463 <COL>

A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17730.1; PID:e125462

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 35.3%; Score 688.5; DB 2; Length 463;
Best Local Similarity 42.8%; Pred. No. 1.3e-33;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 2 VDFGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFGVLPPEINSGRMVAGPGSGPMLAAAAWDGLATELQSTADYGSVSVLT-GVMSGQ 59

QY 62 SAGLMVAAAAPYVAVWMTAQAEELTAQVRAAAAYETAVGLTVPPVIAENRAELMIL 121

DB 60 SSGTMAAAAPYVAVWMTAALAREAAQAASAAAAAYEAFATVPPVVAANRAELAVL 119

QY 122 IATNLLGQNTPAIAVNEAEYEMMAQDAAMFGYAAATATATATATLFPPEAPEMTSAGGL 181

DB 120 AATNIFGQNTGAIAAAEARYAEMMAQDAAMFGYAGSSVAT-QVTPFAAPPPTTNAAGL 178

QY 182 LEQAAVVEASDTAAANQLMNNVPOALQOQAPOGTTPSSKLGGLWKTYS--PHRSP- 238

DB 179 ATQGVAVAAQVAGSAGN-ARSLVSEVLEFLA--TAGTNNKTVASLMNVAITGVVPYASSVY 235

QY 239 -----SNMYSMANNHMTNSGVSMNTNTLSSMLKGFAPAAAQAVQVTAQNGVRA 288

DB 236 NSMLGLGFAESKMWLPANDIVISIFGWFQFQFNFVTFNPDLPK----- 283

QY 289 MSSIGSSLG-----SSGLGG---GVAANLGRAASVGLSVPOQWAAANQAVTFAARALPL 340

DB 284 -SALGAGLGLRLSATSISGLGSTAPASAGSAGSGVGMVPPPSWAAATPAIRTVAAVFSS 342

QY 341 TSLFS--AAERPGQML-----GGLPVGMGARAGGGLSGVLRV 377

DB 343 TGLQVPAAPAISEGLLSQWALASVAGCALGGAAARATGGLGGRV 389

RESULT 13

B70932

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70932

R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70932

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-468 <COL>

A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17730.1; PID:e125462

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 35.0%; Score 682.5; DB 2; Length 468;
Best Local Similarity 40.9%; Pred. No. 2.9e-33;
Matches 164; Conservative 59; Mismatches 125; Indels 53; Gaps 9;

QY 2 VDFGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFGLQPPETISGMYLPGAGPMLAAAVWDGLAELQSMASASIASIVEGMASESWLGP 60

QY 62 SAGLMVAAAAPYVAVWMTAQAEELTAQVRAAAAYETAVGLTVPPVIAENRAELMIL 121

DB 61 SSAGMAAAAPYVVTWMTSAQAAADQAAAVVAYETAFAAVPPVPOIAANRSQILSL 120

QY 122 IATNLLGQNTPAIAVNEAEYEMMAQDAAMFGYAAATATATATLFPPEAPEMTSAGGL 181

DB 121 VATNIFGQNTAIAATEAEYEMMAQDTMAMFGYASSATA-SRLTPTTAPQTTNPSGL 179

QY 182 LEQAAVVEAS-----DTAAANQLMNNVPOALQOQAPOGTTPSSKLGGL- 227

DB 180 AQCAAAATGQATALASGNTNAVTTALSSAAAPFFDIPTLLQGLA--TLSTQYTLQMGQLI 237

QY 228 -----WKTYSPHRSPISNNVSMANNHMTNSGVSMNTNTLSSMLKGFAPAAAQ 276

DB 238 NAIFOPTGATTYQNVFVTAANVTFKSTWANDAMAPNLGMEFKVF-----WQPPPAPE 291

QY 277 AVQTAAQNGVRAMSSGLSSLG-----SSGLGGVAAANLGRAASVGLSVPOQWAAANQAV 331

DB 292 IPK-----SSLGAGLGLRLSAGLSAGLAHAASAGLQANLYGDLSPVPPSWASATPAV 341

QY 332 TPAARALPLTSLTSA-AERGPQMLGGLPVGMGARAGGGL 371

DB 342 RLVTANTPATSLAAAPATQIPANLQGMALGSM---TGGAL 379

RESULT 14

A70646

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70646

R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70646

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-380 <COL>

A:Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PIDN:CA806278.1; PID:e291015; I

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 34.6%; Score 675; DB 2; Length 380;

Best Local Similarity 43.1%; Pred. No. 6.2e-33;

Matches 172; Conservative 56; Mismatches 143; Indels 28; Gaps 12;

QY 2 VDFGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFALLPPEVNSARMYTCGAGSLLLAAAGGWDLSAAELATTAEAYGVSLGLAALHWRGP 60

QY 62 SAGLMVAAAAPYVAVWMTAQAEELTAQVRAAAAYETAVGLTVPPVIAENRAELMIL 121

DB 61 AAEASMAVTAAPYIGWLYTTAEKTOQTATQARAALAFEQAYAMTLPPPVVAANRIQLAL 120

QY 122 IATNLLGQNTPAIAVNEAEYEMMAQDAAMFGYAAATATATLFPPEAPEMTSAGGL 181

DB 121 IATNFFGQNTAIAATEAEYEMMAQDAAMFGYATASA-AAALLTPSPPRQTTNPAGL 179

QY 182 LEQAAVVEASDTAAANQLMNNVPOALQOQAPOGTTPSSKLGGLWKTYSVPHRSP- 238

DB 180 TAQAASVQATDPLSL--LIETVTQALQALTIPFIPEDFTFLDPAIFAGYATVGTQDVE 237

QY 239 SNMYSM--ANNHMSMTNSGVSMNTNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSS 296

DB 238 SFVAGTIGAESNLGILNVGDENPAEVTGDFGIGELYSA-----TSPGGVSA-SGAG--- 289

QY 297 GSSGLGGVGAANLGRAASVGLSVPOQWAAANQAVTFAARALPLTSL--TSAAREG-PGQ 353

Search completed: June 22, 2004, 17:24:52
Job time : 11.9397 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 6.50724 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-14
Perfect score: 1949
Sequence: 1 MVDGALPPEINSARMYAGP.....SGVLRVPPRYVMHPHSPAAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1656.5	85.0	396	1 YD61_MYCTU	Q11031 mycobacteri
2	705	36.2	408	1 YS92_MYCTU	Q10813 mycobacteri
3	688.5	35.3	483	1 Y102_MYCTU	O53951 mycobacteri
4	444	22.8	487	1 Y442_MYCTU	P42811 mycobacteri
5	426.5	21.9	443	1 Y878_MYCTU	Q10540 mycobacteri
6	418	21.4	678	1 YF48_MYCTU	Q10778 mycobacteri
7	364.5	18.7	408	1 SRA_MYCLE	Q07297 mycobacteri
8	324.5	16.6	463	1 Y096_MYCTU	Q10892 mycobacteri
9	324	16.5	434	1 YU18_MYCTU	P15000 mycobacteri
10	321.5	16.5	435	1 YU21_MYCTU	O53268 mycobacteri
11	233.5	12.0	178	1 Y125_MYCTU	O06246 mycobacteri
12	217.5	11.2	176	1 Y125_MYCTU	Q50703 mycobacteri
13	153.5	7.9	860	1 ELS_MOUSE	P54320 mus musculu
14	151.5	7.8	232	1 YV26_MYCTU	Q50702 mycobacteri
15	150	7.7	864	1 ELS_RAT	Q99372 rattus norv
16	143.5	7.4	730	1 ELS_HUMAN	P15502 homo sapien
17	139.5	7.2	881	1 PRY3_YEAST	P47033 saccharomyc
18	135	6.9	907	1 A180_HUMAN	O60641 homo sapien
19	133.5	6.8	825	1 ICP0_HSV2H	P28284 herpes simp
20	133.5	6.8	2432	1 Y13R_IRV6	P18305 chilo iride
21	132	6.8	1120	1 STFR_ECOLI	P76072 escherichia
22	132	6.8	2090	1 N214_HUMAN	P35658 homo sapien
23	131	6.7	836	1 VQ26_BPMU5	O05233 mycobacteri
24	131	6.7	1150	1 APMU_PIG	P12021 sus scrofa
25	129	6.6	1783	1 RAA3_CHLRE	Q9164 chlamydomon
26	128	6.6	779	1 SRP_CHROME	P52172 drosophila
27	127.5	6.5	790	1 ANP_NOTCO	P24856 nototheria
28	126	6.5	354	1 YAU3_SCHPO	Q10169 schizosacch
29	125.5	6.4	1211	1 BUN2_DROME	Q24523 drosophila
30	125	6.4	577	1 GST2_HUMAN	P33240 homo sapien
31	124	6.4	2090	1 HFCL_MESAU	P51611 mesocricetu
32	123.5	6.3	677	1 Y136_MYCTU	Q50597 mycobacteri
33	123	6.3	432	1 YF10_MYCTU	P71789 mycobacteri

ALIGNMENTS

RESULT 1

ID	YD61_MYCTU	STANDARD	PRT	396 AA
AC	Q11031;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical PPE-family protein RV1361C/MT1406.			
GN	RV1361C OR MT1406 OR MYC02B10.25C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Biglieri K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Taylor K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RX	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey B.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	J. Bacteriol. 184:5479-5490(2002).			
CC	!- SIMILARITY: Belongs to the mycobacterial PPE family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z75555; CAA99966.1; -.			
DR	EMBL; ABC07013; AAK45669.1; -.			
DR	PIR; H70741; H70741.			
DR	TIGR; MT1406; -.			
DR	Tuberculist; RV1361C; -.			
DR	InterPro; IPR000030; Microbac_PPE.			

34	123	6.3	1140	1 YM96_YEAST	Q04893 saccharomyc
35	122	6.3	394	1 HYF1_ALCEU	P45805 alcaligenes
36	121.5	6.2	1025	1 SLAP_CAUCR	P35828 caulobacter
37	121.5	6.2	1845	1 Z236_HUMAN	Q8L36 homo sapien
38	121	6.2	1199	1 P121_RAT	P28591 rattus norv
39	119.5	6.1	635	1 HMLA_DROME	P10105 drosophila
40	119	6.1	915	1 A180_RAT	Q05140 rattus norv
41	118.5	6.1	580	1 EXPR_XANCP	P23314 xanthomonas
42	118.5	6.1	2038	1 FSH_DROME	P13709 drosophila
43	118	6.1	1508	1 BCSC_XANAC	P88938 xanthomonas
44	117	6.0	444	1 Y808_CHLPN	Q87798 chlamydia p
45	117	6.0	652	1 PICA_HUMAN	Q13492 homo sapien

```

DR Pfam: PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 158 159 TA -> AT (IN REF. 2).
SQ SEQUENCE 396 AA; 40015 MW; 6AFAE0D7B5F668D0 CRC64;

Query Match      85.0%; Score 1656.5; DB 1; Length 396;
Best Local Similarity 85.1%; Pred. No. 1.3e-89;
Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPEINSMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFCAATATATATATLPPFEAPEMTSAGG 180
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFCAATATATATLPPFEAPEMTSAGG 180
QY 181 LLEQAAVEASDTAAANQLMNVPQALQAQTOGTTFSSKLGGLWKTVPVHRSPISN 240
DB 181 LLEQAAVEASDTAAANQLMNVPQALQAQTOGTTFSSKLGGLWKTVPVHRSPISN 240
QY 241 MVSMAHNMSTNGSVMTNTLSSMLKGFAPAAAAQVTAQAQNGVAMSS----LGSSL 296
DB 241 IVSMNNHVSMTNGSVMASTLHSMKGFAP-AAAQAVETAQAQNGVAMSSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAARPGQMIG 356
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAARPGHMIG 359
QY 357 GLPVGQMGARAG--GGISGLVRVPPVYVMPHSPAG 391
DB 360 GLPLGQLTSGGGGVSNALRMPPRAYVMPVPAAG 396

RESULT 2
YS92 MYCTU STANDARD; PRT; 408 AA.
ID Q10813;
AC Q10813;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV2892C/Mt2959/Mb2916c.
GN RV2892C OR MT2959 OR MFCY274.23C OR MB2916C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RN [2]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544 (1998).
[2]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

```

```

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthey S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: Z74024; CAA98377.1; -
CC EMBL: AE071119; AAK47285.1; -
CC EMBL: EX248344; CAD96603.1; -
CC FIR: G70925; G70925.
CC TIGR: MT2959; -
CC Tuberculist; RV2892C; -
CC InterPro; IPR00030; Microbac_PPE.
CC Pfam: PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 56 76 POTENTIAL.
SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F20D7827199 CRC64;

Query Match      36.2%; Score 705; DB 1; Length 408;
Best Local Similarity 41.8%; Pred. No. 2.7e-34;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

QY 2 VDFGALPPEINSMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIGS 61
DB 1 MDFGVLPEINSGMYAGPGSAPMMAAAWDSLAELGLAAGGYRLAISLTGAYWAGP 60
QY 62 SAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 121
DB 61 AAASMVAATPYVAVLSATAGQAQAGQAARAAAAYELAFAMTVPPVIVANRALLVAL 120
QY 122 IATNLGQNTPAIAVNEAEYGEWMAQDAAMFCAATATATATLPPFEAPEMTSAGGL 181
DB 121 VATNFFGQNTPAIAATEAQVAEMWQAQDAAMAYAGSAAIAT-ELTPTTAAPVTTSPAAL 179
QY 182 LEQAAA-----VEEASDTAAANQLMN-----VP--QAQQ--LAQQTGTTTSSKL--- 224
DB 180 AGQAAATVSVSTVPPLATTAAPVQLQLQSSTSLIPWYSALQQWLAENLLGUTPNRMTIV 239
QY 225 -----GGLWKTVPVHRSPISNMVSMNNHMTNNGSVMTNTLSSMLKGFAPAAAA 275
DB 240 RLIGISYFDEGL-----LQFEASLAQAQAPGTPGGAG--DSGSSVLDMSGTIFA 287
QY 276 QAVQTAQNGVAMSSL--GSSLSG-----SGLGGVAAANLGRAASVGSLS 319
DB 288 -----GPRASPSVAGGAVGVQTPQYVWYWDRESIGGSVSAALKGSSAGLS 338
QY 320 VPQAWAANQAVTPAARALP---LTSITSAERPGQMLGGLPVGQMGARAGGSLGVLR 376
DB 339 VPDWAARARWANPAANWELPGDDVTALRGTAENA--LLRGFPWASAGQTGGGF--VHK 393
QY 377 VPPRPYVMPHSPAG 391
DB 394 YGFRFLAVNQRPFPAG 408

```

```
RESULT 3
YI02 MYCTU STANDARD; PRT; 463 AA.
AC O53951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV1802/MT1851/MD1830.
GN RV1802 OR MT1851 OR MTV049.24 OR MB1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kelton J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterlinden T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
SEQUENCE FROM N.A.
RP SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Mounsepe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -I- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AL022021; CAAL1723.1; -
CC ENBL; AE007044; AAK46123.1; -
CC ENBL; BX248340; CAD94533.1; -
CC PIR; C70931; C70931.
CC TIGR; MT1851; -
CC TubercuList; RV1802; -
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 401 401 S -> I (IN REF. 2).
```

```
SQ SEQUENCE 463 AA; 46021 MW; EE64828BFO9FA551 CRC64;
Query Match 35.3%; Score 688.5; DB 1; Length 463;
Best Local Similarity 42.8%; Pred. No. 2.8e-33;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 2 VDFGALPPEINSEARMYAGPGSASLVAAQWDSVADSLFAASAFQSVWGLTVGWSIGS 61
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MDFGVLPPEINSEGRMYAGPGSGPMLAAAAAAMDGLATELQSTAADYGSVISVLT-GWMSGQ 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 SAGLMVAAAAPYVAVMSVTGAQELTAQVRVAAAAYETAYGLTVPPVTAENPAELMIL 121
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 SSGTMAAAAAAPYVAVMSATLAAREAAAAQASAAAAAYEAFATVPPVVAANRAELAVL 119
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLFFEPAPETSAGGL 181
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 AATNIFGONTGAIATAAEYAEWMAQDAAMFGYAGSSVAT-QVTFFAAPPTTNAAGL 178
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 182 LEQAAVEEASDTAAANQLMNVPAQLQLAQTPQGTTPSKLGLLWKTYS--PHRSP1- 238
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 ATQGVAVAAQAVGASAGN-ARSLNSEVLEFLA--TAGTNYNKTVASLMNAVTVGVPIASSVY 235
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 239 -----SNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRA 288
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 236 NSMTGLGFAESKMWLPANDTVISTIFGMVQFQKEFNVTFNPDLPK----- 283
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 289 MSSLGSSLG-----SSGLGG--GVAANLGRPAASVGLSVDPQAAWAAANQAVTPAARALPL 340
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 -SALGAGLGLRSATSSGLGSTAPASAGASQAGSVGMSVPPSWAAATPAIKTVAAVFSS 342
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 341 TSLTS--AAERGPQGM-----GGLPYGQMGARAGGSLSGVLRV 377
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 343 TGLQVPAAPAAISEGSLLSQVALASVAGGALGAAARATGGLGGRV 389
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
Y442 MYCTU STANDARD; PRT; 487 AA.
ID Y442 MYCTU
AC P42611; OS3727;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0442c/MT0458.
GN RV0442C OR MT0458 OR MTV037.05C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Erdmann;
RX MEDLINE=87137260; PubMed=3029018;
RA Shinnick T.M.;
RT "The 65-kilodalton antigen of Mycobacterium tuberculosis."
RL J. Bacteriol. 169:1080-1088 (1987).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
```

```

RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15467; AAA88235.1; ALT INIT.
DR EMBL; AL021932; CAA17399.1; -.
DR EMBL; AE006948; AAK44681.1; -.
DR PIR; C70830; C70830.
DR TIGR; MT0458; -.
DR TubercuList; Rv0442c; -.
DR InterPro; IPR00030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 5.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 40 40 E -> K (IN REF. 2).
FT CONFLICT 96 96 I -> T (IN REF. 1).
FT CONFLICT 211 211 G -> GNNIG (IN REF. 1).
FT CONFLICT 487 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;
SQ SEQUENCE 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;

Query Match 22.8%; Score 444; DB 1; Length 487;
Best Local Similarity 32.1%; Pred. No. 4.8e-19; Indels 58; Gaps 12;
Matches 135; Conservative 48; Mismatches 180;

Qy 4 FGALPPPEINGARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGSSA 63
Db 6 FAWLPPEINGALMPAGPGSGLPIAAATAWGELAEELLASISLGSVTSLSLTSGLWLGPSA 65
Qy 64 GLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMILIA 123
Db 66 AAMMAVATQYLAWLSTAAQAEQAQAQAQAIAATAFEAALAAATVQPAVAAANRGLMQLAA 125
Qy 124 TNLIGQNTPAIVNNEAEYGEWMAQDAAMFGYAAATATATATLTPPEAPMTSAG--- 179
Db 126 TWFQGNAPALMDVEAAEQWALDVAAVAGYHFDASAAVAQLAPHQOV--LRNLGIDIG 183
Qy 180 --GLLEQAAVAVERASDTAAANQLMNNVPQALQOLAQPTQCTTSSKLG-----GLW 228
Db 184 KNGQINLGFNGTSGNIGNNNIGNNIGNTGTGNTGNTGNTGNTGNTGNTGNTGNTGNTG 243
Qy 229 KTVS-----PHRSPISNMVSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVQ 279
Db 244 NTSGNIGFGITGDHQMGFGFGNFGSGN-IGFNGSNTGNVGLFNS----- 287
Qy 280 TAAQNGVRAMSLGSSIGSLGGVAAANLGRAASVGSLSVPQAAANAAQVTPAARALP 339
Db 288 GSGNIGTSGNSLNGTSGT---INAGLG---SAGSLNT-SFWVAGNQNAALGSAAGS 340
Qy 340 LTSITSAAERGPQM-----LGLPLVQMGARAG--GGLSGVLR--VPRPVPYMPHPSPA 390
Db 341 EALVSSAGYATGGMSTAAALSSGILASLSTGGLQHLANVLNLSLTNTPTVAAPASAV 400
Qy 391 G 391
Db 401 G 401

```

```

Y878 MYCTU STANDARD; PRT; 443 AA.
ID Y878 MYCTU
AC Q10540;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv0878c/MT0901.
GN Rv0878c OR MT0901 OR MTC331.06c.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1] RN
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2] RN
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol 184:5479-5490(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73101; CAA97385.1; -.
DR EMBL; AE006977; AAK45143.1; ALT_INIT.
DR PIR; C70780; C70780.
DR TIGR; MT0901; -.
DR TubercuList; Rv0878c; -.
DR InterPro; IPR00030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 4.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT DOMAIN 64 73 POLY-ALA.
FT DOMAIN 81 115 ALA-RICH.
FT DOMAIN 231 270 4 X 10 AA APPROXIMATE REPEATS.
SQ SEQUENCE 443 AA; 43592 MW; C58BEC607F0675E2 CRC64;

Query Match 21.9%; Score 426.5; DB 1; Length 443;
Best Local Similarity 31.2%; Pred. No. 4.5e-18; Indels 87; Gaps 14;
Matches 125; Conservative 56; Mismatches 132;

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entitled requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z74020; CRA98335.1; -
DR EMBL; AE007026; AAK45866.1; ALT_INIT.
DR PIR; A70762; A70762.
DR TIGR; MT1599; -
DR Tuberculin; RV1548C; -
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 11.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT CONFLICT 258 258 D -> G (IN REF. 2). CRC64;
SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;

Query Match 21.4%; Score 418; DB 1; Length 678;
Best Local Similarity 31.9%; Pred. No. 2.2e-17;
Matches 106; Conservative 48; Mismatches 140; Indels 38; Gaps 7;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAPQSVTWGLTVGSWIGS 61
DB 1 MNFVLPPEVNSARIYAGAGPAPMLAAAVANDCLAEELGWAASFSLLISGLTAGPGSAW 60
QY 59 IGSSAGLVAAASPYVAMSVTVTAQAEELTAQVRVAAAAYETAYGLTPPPVIAENRAEL 118
DB 61 QGPAAMAAAAPYLSWLNATARAEGAAAGAKAAAVYEAARAATAHPALVAANRNOL 120
QY 119 MILIATNLGONTPATVNEAEYGEWGAODAAAMFCYAAATATATATLLPPEEAPMTSA 178
DB 121 LSLVNLNFGQNLPAALAAEASTEQIWAQVAAWGYHGGASTVASQLTPWQQ----- 173
QY 179 GGLEQAAAEVEASDTAAANQLMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 238
DB 174 -LLSVLPVVTAPAGAV-----GVPA--LAIPALGV--ENIG-----V 208
QY 239 SNMV-----SVANEMSMTNSG-----VSMTN-----TLSSMLGFPAPAAAQAV 278
DB 209 GNFGLGIGNNVGSGNTGDNFYNGFGIGNANLNGNTGNANLGSNGAGFFNFGNGNDG 268
QY 279 QTAQNGVRAMSLGS-SLGSGLGGVAA--NLGRAASVGSLSVPOAAANQAVTPAA 335
DB 269 NTNFGSGNAGFLNIGSGNEGSLNFGNAGDDNTG-----WNSGD----- 309
QY 336 RALPLTSLTSAABRGQOMLGGI--PVGQMGAPAGGSLG 373
DB 310 -----INTGFNSGDLNTGIGSPVTQGVANSFGNTG 341

RESULT 6
YF48_MYCTU STANDARD; PRT; 678 AA.
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV1548C/MT1599.
GN RV1548C OR MT1599 OR MCY48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., Nelson W.C., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490 (2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

RESULT 7

SRA_MYCLE STANDARD; PRT; 408 AA.
ID Q07297;
AC Q07297;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine-rich antigen (25L) (45 kDa protein).
GN SRA OR ML0411 OR MLCU383.14.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93239328; PubMed=8478104;
RA Vega-Lopez F., Brooks L.A., Dockrell H.M., de Smet K.A.L.,
RA Thompson J.K., Hussain R., Stoker N.G.;
RT "Sequence and immunological characterization of a serine-rich antigen
RT from Mycobacterium leprae";
RL Infect. Immun. 61:2145-2153 (1993).
RN [2]
RP SEQUENCE FROM N.A.

RX MEDLINE=95020554; PubMed=7934845;
 RA Rinke de Wit T.F., Clark-Curtiss J.E., Abebe F., Kolkon A.H.J.,
 RA Jonson A.A.M., Thole J.E.R.;
 RT "A Mycobacterium leprae-specific gene encoding an immunologically
 RT recognized 45 kDa protein."
 RL Mol. Microbiol. 10:829-838(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser J., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 RC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U00015; AAC43220.1; -
 DR EMBL; X68431; CAA48480.1; -
 DR EMBL; Z21952; CAA79950.1; -
 DR EMBL; Z97179; CAB09938.1; -
 DR EMBL; AL583918; CAC29919.1; -
 DR PIR; C86960; C86960.
 DR PIR; S33522; S33522.
 DR PIR; S39872; S39872.
 DR LepPro; ML0411; -
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Antigen; Repeat; Complete proteome.
 FT DOMAIN 192 196 POLY-SER.
 FT DOMAIN 209 235 2 X 6 AA REPEATS OF S-V-A-Q-S-E.
 FT REPEAT 209 214 1.
 FT REPEAT 230 235 2.
 FT CONFLICT 132 132 T -> S (IN REF. 2).
 FT CONFLICT 189 189 S -> L (IN REF. 2).
 FT CONFLICT 191 191 H -> D (IN REF. 2).
 FT CONFLICT 292 292 P -> L (IN REF. 2).
 SQ SEQUENCE 408 AA; 42466 MW; 5C0C2BE0D6BEA9D8 CRC64;
 Query Match 16.7%; Score 364.5; DB 1; Length 408;
 Best Local Similarity 26.3%; Pred. No. 1.7e-14;
 Matches 115; Conservative 72; Mismatches 163; Indels 87; Gaps 11;
 Qy 1 MVDGALPEINSRMVYAGSGSLVAQAQWDSVASDLFSASAFQSVWGLTVGSWIG 60
 Db 1 MFDPMVYSPEVNAFLMSRGPGTSLWGAARWISLAQLMENAQEVSDTIWVAVPASFG 60
 Qy 61 SSAGLMAAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTVPVPIANRAELMI 120
 Db 61 ETSMDLASRSTVFVWLDGNAENAGLTARVLAHVAVAFEEARAGMVPLLTVLGNIIHTWA 120
 Qy 121 LIATNLGQNTPAJANVEAYGEMWQAQAAAPGYAAATATATATLLPPEAPEMTSAGG 180
 Db 121 LKAINWFGQVSTTVAALEADYDLMMVQNSTANTTYRDLVLRGTGKMFEPAPQLVSR-Y 179
 Qy 181 LLEQAAAEASDSTAANQALMNNVPAQLQALQ-
 Db 180 CMERDSVNSFHSSSSSDSLYESIDNLYDSVAQSEEHGSDMSQSYNTCGSVAQSELCD 239

QY 214 ----PTQ-----GTTPTSSKLGGLWKTIVSPHRSPISNMVSMANNHMTNNGSVMTNT 261
 Db 240 PGTFSQSSQNDLSATSLTQQLGGL-----DSIISASASALLTINS--ISSST 286
 QY 262 LSSMLKGFAPAAAAQAVQTAQNGVYRAMSSLGSSLGSSGLGGGVAAN-----LGRASAVG 316
 Db 287 ASSIM-----PIVASQVTTGLRSQV-AVEKMIQS-SSTAVSVDVAASKVAVGQAVSVG 341
 QY 317 SLSVPCWAAANQAVTPAARALP--LTSILTSAAERPGQMLGGLPVGQWGAAGGGLSGV 374
 Db 342 ALRVENWATASQPVNATAHSPAGCSAITTA-----VSGPLEGV 381
 QY 375 LRVPPRPVYVMPHSPAAG 391
 Db 382 TQ--PAEEVLTASVAGG 396
 RESULT 8
 Y096.MYCTU
 ID Y096 MYCTU STANDARD; PRY 463 AA.
 AC Q10892;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Hypothetical PPE-family protein RV0096/MT0105.
 GN RV0096 OR MT0105 OR MTCY251.15.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_Taxid=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z74410; CAA98932.1; -
 DR EMBL; AE006922; AAK44327.1; -
 DR PIR; H70750; H70750.
 DR TIGR; MT0105; -
 DR TubercuList; RV0096; -


```
DR InterPro; IPR000030; Microbac_PPE.
KW Pfam; PF00823; PPE; 1.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
SQ SEQUENCE 463 AA; 429D66A033D0D8 CRC64;

Query Match
Best Local Similarity 16.6%; Score 324.5; DB 1; Length 463;
Matches 112; Conservative 54; Mismatches 144; Indels 93; Gaps 12;

QY 6 ALPPEINARMYAGPGSGASLVAAQWMDSVASDLFGAASAFQSVWGLTVGSGWIGSAGL 65
DB 2 AIPEVHSGLLSAGCGPGSLVAAQWQELSDQYALACALGQLLGEVQASSWQGTAAQ 61
QY 66 MYAASPYVAMSVTAGQAEILTAQVRVAAAYETAYGLTVPPVIAENRAELMILIATN 125
DB 62 YVAAGPYLAWLEQTAINSVTAQHVAAAYCSALAAFTPAELAAHAIHGVLIATN 121
QY 126 LLGONTPTAIVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPEMTSAGLLLEQA 185
DB 122 FFGINTVPIALNEADYVRMQLQADTNAVQVADAATVAVPTQAPPIRAPGG----- 176
QY 186 AAVEEASDT-----AAANQLMNVPQALQOLAQFTQ-----GTPSSKL----- 224
DB 177 ----DAADTRLVLSIGLIRDI---LDFIANPYKVFLEFFQFGFSPAVTVVLVALVAL 229
QY 225 ---GGLW-----KTVSPHRSPINMWSMANNHSMNTNSGYSMNTLSSMLK-----GF 269
DB 230 QLVDFLWYPYVYAGVGLLLPFFTP-----TLSAUTALSALHLLNLPAGL 275
QY 270 AAAAAQAVQTAQNGVRAMSSLSGSLGGVAAANLGRAASVGSLSVPOWAAANQ 329
DB 276 LPIAAA-----LPGDQWGANLAVATATAVP-----GSGP 308
QY 330 AVTPAPRALPLTSLTSAERGGP---QMLGGLPVG-OMGARAG 368
DB 309 PTSPAPAPSPSSVSGSAAAPGISYAVPLAPPVSGSPKAG 351

RESULT 9
YU18_MYCTU STANDARD; PRT; 434 AA.
AC P31500; O53265;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3018C/MT3098/MT3101.
GN RV3018C OR MT3098/MT3101 OR MTV012.32C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
```

```

Db 298 GGPLLGAARAAVPGVAGLAGVLAAL-PAVGAA--AGAPAAALVGSVPVSGVSPQA 354
QY 324 WAAANQAVTPAARALPILTSLSAERPGQMLGSLPVGQWAGARAGGSLGV 374
Db 355 RLVS--AVEPAPASTSVSL--ASDRGAGAL--GF-VGTAGKESVGGPAGL 398

RESULT 10
YU21 MYCTU
ID YU21 MYCTU STANDARD; PRT; 435 AA.
AC 053268; OS3269;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3021C/RV3022C/MT3106.
GN RV3021C/RV3022C OR MT3106 OR MTV012.35C/MTV012.36C.
OS Mycobacterium tuberculosis
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 82.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021287; CAAL16106.1; ALT_FRAME.
DR EMBL; AL021287; CAAL16107.1; ALT_FRAME.
DR EMBL; AE007129; AAK47435.1; -.
DR TIGR; MT3106; -.
DR TubercuList; RV3021C; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
KW CONFLICT 299 G -> A (IN REF. 2).
FT CONFLICT 317 LAGV -> VTGL (IN REF. 2).
FT CONFLICT 326 L -> V (IN REF. 2).
SQ SEQUENCE 435 AA; 42876 MW; 38157643EAA8484A CRC64;

```

```

Query Match 16.5%; Score 321.5; DB 1; Length 435;
Best Local Similarity 26.5%; Pred. No. 5.6e-12;
Matches 115; Conservative 53; Mismatches 15; Indels 107; Gaps 12;

QY 6 ALPPPEINARMVAGPGSASLVAAQAQMDSDVSLFSAASAFQSVVWGLTVSGWIGSSAGL 65
Db 8 ASPPEVHSALLSAGPGPSLQAAAAGWSALSAAEVAQAQELSVAQVAVAGAGVWQGPSAEL 67
QY 66 MYAAASPYVAMVSVTGAQELTAQAQVRVAAAYTAYGLTVPPPPVIAENRAELMILITATN 125
Db 68 FVAAYVPVWMLVQASADSAASAAAGHEHAAGYVCAALAEEMPTLPALAANHLTHAVLVATN 127
QY 126 LLGQNTPAIVNEAYGEMWQAQDAAMAFGYAA----- 157
Db 128 FFGINTIPIALNEADYVRWVQAATVMSAYEAVVGAALVATPHGTGPAPVIVKPGANEASN 187
QY 158 AVATATATALLPPEE-----APEMTSAGGLLEQAAAEEASDTAAANQIMNNVPOALQ 209
Db 188 AVAATATIPFPFGLAKLEMAAQAFTEVGLIMKSAEAWAVGFVELITGLVNFEP---- 243
QY 210 QLAQPTQGTTPSSKLGGLWKTIVSPHRSPISNVSM-----ANNH 248
Db 244 -----WLW-----LTGMIDMFFATVGFALGVFLVLLLEFAVVLE 278
QY 249 MSMTNSGVSMNTLSS-----MLKGFAPAAAQAQVQTAQAQNGVRAVMSISLSSGLSSGL 301
Db 279 LAILSIGHIISNFGAIPVLGGPLLGAALAAVPGVAGLAGVAGLAALPAVGAAGAP-- 336
QY 302 GCGVAANTGRAASV--GSLSVPOQMAAANQAVTPAARALPLTSLTSAARPGQMLGLPLV 360
Db 337 ----AALVGSVAPVSGGVVSPQARLVS--AVEPAPASTSVSL--ASDRGAGAL--GF-V 385
QY 361 GQMGARAGGSLGV 374
Db 386 GTAGKESVGGPAGL 399

RESULT 11
YU29 MYCTU
ID YU29 MYCTU STANDARD; PRT; 178 AA.
AC 006246;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3429/MT3533.
GN RV3429 OR MT3533 OR MTCV77.01.
OS Mycobacterium tuberculosis
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 82.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021287; CAAL16106.1; ALT_FRAME.
DR EMBL; AL021287; CAAL16107.1; ALT_FRAME.
DR EMBL; AE007129; AAK47435.1; -.
DR TIGR; MT3106; -.
DR TubercuList; RV3021C; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
KW CONFLICT 299 G -> A (IN REF. 2).
FT CONFLICT 317 LAGV -> VTGL (IN REF. 2).
FT CONFLICT 326 L -> V (IN REF. 2).
SQ SEQUENCE 435 AA; 42876 MW; 38157643EAA8484A CRC64;

```

```
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RL laboratory strains";  
CC J. Bacteriol. 184:15479-5490(2002).  
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; Z95389; CAB08678.1; -;  
DR EMBL; AE007158; AAK47873.1; -;  
DR PIR; C70975; C70975.  
DR TIGR; MT3533; -;  
DR TubercuList; RV3429; -;  
DR InterPro; IPR000030; Microbac_PPE.  
DR Pfam; PF00823; PPE; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 176 AA; 19811 MW; 8BE1FC025ABFBEA6 CRC64;  
  
Query Match 12.0%; Score 233.5; DB 1; Length 178;  
Best Local Similarity 35.9%; Pred. NO. 2.8e-07;  
Matches 65; Conservative 25; Mismatches 76; Indels 15; Gaps 3;  
  
QY 7 LPPEINARMYAGPGSASLVAAQAQ----NWDS---VASDLFAASAFQSVVWGLTVGSMI 59  
DB 5 IPAEYISNIIEYEGPGADSLFAAQQLRLMYNANMTAKSLTRDLGELQF-----NWK 56  
QY 60 GSGAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVFAENRAELM 119  
DB 57 GSSSLMDADAAGRYLDWTKHRSQILETAYVIDFLAYVVEETRHKKVVPATIANNEEVH 116  
QY 120 ILTATNLLQNTPAIAVNEAEYGEWQAQDAAMFGYAAATATATATATLLPFEAPMTSAG 179  
DB 117 RLFAISVAGVNTPAIAGLDAQYQYRAQNAIVNNDYQSTARFILAYLPWQEPPIYGGG 176  
QY 180 G 180  
DB 177 G 177  
  
RESULT 12  
Y25 MYCTU STANDARD; PRT; 176 AA.  
AC Q50703;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical PPE-family protein RV3425.  
GN RV3425 OR MTCY78.04C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterinae; Mycobacteriaceae; Mycobacterium.  
OX NCBI_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,  
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence";  
RL Nature 393:537-544(1998).
```

```
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; Z77165; CAB01031.1; -;  
DR PIR; F70738; F70738.  
DR TubercuList; RV3425; -;  
DR InterPro; IPR000030; Microbac_PPE.  
DR Pfam; PF00823; PPE; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 176 AA; 19855 MW; B8CFE2E9463B87B0 CRC64;  
  
Query Match 11.2%; Score 217.5; DB 1; Length 176;  
Best Local Similarity 33.5%; Pred. NO. 2.3e-06;  
Matches 58; Conservative 31; Mismatches 83; Indels 1; Gaps 1;  
  
QY 7 LPPEINARMYAGPGSASLVAAQAQWDSVADLFAASAFQSVVWGLTVGSGWSSAGLM 66  
DB 5 IPAEYISNIIEYEGPGADSLFFASQQLRELAYSVETTAESLEDELDELDEL-ENWKSSSDDL 63  
QY 67 VAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVFAENRAELMILATNL 126  
DB 64 ADAVERYLQWLSKSSQLKHAAWVINGLANAYNDRKRWVPEEIAANRERRRLIASNV 123  
QY 127 LGQNTPAIAVNEAEYGEWQAQDAAMFGYAAATATATATATLLPFEAPMTSAG 179  
DB 124 AGVNTPAIADLDAQYDQYRARNVAMNAYVSWTRSAUSDUPRWRPPPIYRGG 176  
  
RESULT 13  
ELS_MOUSE STANDARD; PRT; 860 AA.  
AC PS4320;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Elastin precursor (Tropoelastin).  
GN ELN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c; TISSUE=Lung;  
RX MEDLINE=95130069; PubMed=7829060;  
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;  
RT "Use of an intron polymorphism to localize the tropoelastin gene to  
RT mouse chromosome 5 in a region of linkage conservation with human  
RT chromosome 7";  
RL Genomics 23:125-131(1994).  
CC -!- FUNCTION: Major structural protein of tissues such as aorta and  
CC nuchal ligament, which must expand rapidly and recover completely.  
CC -!- SUBUNIT: The polymeric elastin chains are cross-linked together  
CC into an extensible 3D network.  
CC -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.  
CC -!- PTM: The crosslinks are made of deaminated Lys.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; U08210; AAA80155.1; -;
```


CC -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=8;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=1;
CC Name=2;
CC Name=3;
CC Name=4;
CC Name=5;
CC Name=6;
CC Name=7;
CC Name=8;
CC IsoId=Q99372-1; Sequence=VSP_004244;
CC IsoId=Q99372-2; Sequence=VSP_004245;
CC IsoId=Q99372-3; Sequence=VSP_004246;
CC IsoId=Q99372-4; Sequence=VSP_004247;
CC IsoId=Q99372-5; Sequence=VSP_004248;
CC IsoId=Q99372-6; Sequence=VSP_004249;
CC IsoId=Q99372-7; Sequence=VSP_004250;
CC IsoId=Q99372-8; Sequence=VSP_004251;
CC -!- PTM: The crosslinks are made of deaminated Lys.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M60647; AAA42269.1; -;
CC EMBL; J04035; AAA42268.1; -;
CC EMBL; M86372; AAA42271.1; -;
CC EMBL; M86355; AAA42271.1; JOINED.
CC EMBL; M86363; AAA42271.1; JOINED.
CC EMBL; M86364; AAA42271.1; JOINED.
CC EMBL; M86366; AAA42271.1; JOINED.
CC EMBL; M86371; AAA42271.1; JOINED.
CC EMBL; M86376; AAA42272.1; -;
CC EMBL; M86373; AAA42272.1; JOINED.
CC EMBL; M86375; AAA42272.1; JOINED.
CC PIR; A36106; EART
CC InterPro; IPR003979; tropoelastin.
CC RefSeq; PRO1500; TROPELASTIN.
CC Structural protein; Connective tissue; Repeat; Signal;
CC Alternative splicing.
CC NON TER 1
CC SIGNAL <1 21 BY SIMILARITY.
CC CHAIN 22 864 ELASTIN
CC DISULFID 854 859 BY SIMILARITY.
CC VARSPLIC 263 307 Missing (in isoform 2, isoform 5, isoform
7 and isoform 8).
FT VARSPLIC 308 308 /FTid=VSP_004244.
FT VARSPLIC 308 308 Missing (in isoform 3, isoform 5, isoform
6 and isoform 8).
FT VARSPLIC 809 823 /FTid=VSP_004245.
FT VARSPLIC 809 823 Missing (in isoform 4, isoform 6, isoform
7 and isoform 8).
SQ SEQUENCE 864 AA; 72786 MW; 456894BE09E79FD4 CRC64;

Query Match 7.7%; Score 150; DB 1; Length 864;
Best Local Similarity 25.9%; Pred. No. 0.11;
Matches 102; Conservative 31; Mismatches 153; Indels 108; Gaps 18;

QY 5 GALP-----PBIAGARMYAGPGSASLVAAQMDVSASDIFSAASAFQSVVMGLTVGSW-- 58
Db 399 GGIPGVGGPGIGGPGVGGVSPAAAKAAKAAK--YGARGGVGIPTYGVGAGGFP 456

QY 59 --TGSSAGLMAASPVAMWSTAGQAELETAQVRVAAYETAYGLTVPPPIAENRA 116
Db 457 YGVGAGAGL-----GGASQAAAA--AAAAKAAKYAG----- 486

QY 117 ELMILINLLGQNTPAIAVNEAEYGEWMAODAAAMFGYAAATATATATILPPEAPEMT 176
Db 487 -----GAGTLGGLVPG-----AVFGALPGAVFGALPGAVPGALP-GAVFGVP 527
QY 177 SAGGLL---EQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSS-KLGLWKTIVS 232
Db 528 GTGGVPGAGTPAAAAAATAAATAA-----KAGQYGLGPGVGGVGGVGGVGGV 579
QY 233 PHSPISNMYSMANNHMTNSGVSM-TNTLSSMLKGF-APAAAAQAVQTAQAQNGVRAMS 290
Db 580 PG-----GVTGIGTGPCTGLVPGDLGGAGTPAAAKSAKAAKAAKAYRAAA 624
QY 291 SLGSSLSGSSGLGGV-----AANLGRAASV---GSLSVPOAWAANAQAVTPAARALPLT 341
Db 625 GLGAGVPLGVGAGVPGFAGAGGAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGV 681
QY 342 SLTSAARERGPQOMLGGLPVQMGARAG-GGLSGV 374
Db 682 -----GGPGGLGGPGGLGGPGGPGGPGGLGGV 708

Search completed: June 22, 2004, 17:16:54
Job time : 7.59814 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 31.9703 Seconds

(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349A-14

Perfect score: 1949

Sequence: 1 MVDFGALPEINSARMYAGP.....SGVLVPPRPYVWPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioplasmid:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1949	100.0	391	16	005298
2	1920.5	98.5	390	16	Q7U0E9
3	1779.5	91.3	396	16	Q7U071
4	1648	84.6	393	16	Q8VIZ3
5	1583	81.2	393	16	Q06341
6	1583	81.2	393	16	Q7TW55
7	779.5	40.0	393	16	Q7TZJ3
8	779.5	40.0	411	16	Q53939
9	754	38.7	410	2	Q99011
10	751	38.5	409	16	Q53957
11	749	38.4	409	16	Q7TZH7
12	740.5	38.0	399	16	Q7TZH8
13	737.5	37.8	403	16	Q53956
14	736	37.8	423	16	Q53950
15	736	37.8	423	16	Q7TZI4
16	731	37.5	421	16	Q9Z5K0

17	702	36.0	391	16	P96362
18	702	36.0	391	16	Q7U0T5
19	702	36.0	413	16	Q06386
20	683.5	35.1	694	16	Q8VUW0
21	682.5	35.0	468	16	O53958
22	675	34.6	380	16	P95190
23	675	34.6	380	16	Q7TX66
24	670.5	34.4	385	16	Q7TZ87
25	669.5	34.4	394	16	Q7TXX5
26	668.5	34.3	462	16	Q3310
27	667.5	34.2	385	16	Q33204
28	666.5	34.2	385	16	Q8VJZ0
29	641	32.9	364	16	Q7TZJ5
30	637	32.7	350	16	Q7TZJ2
31	635	32.6	363	16	O53940
32	633.5	32.5	365	16	O86373
33	629.5	32.3	405	16	Q8VJW5
34	626.5	32.1	381	16	Q7TX67
35	618.5	31.7	397	2	Q9AGF0
36	615.5	31.6	382	16	Q7TXX3
37	615.5	31.6	402	16	Q33312
38	610	31.3	443	16	Q8VKL9
39	604	31.0	443	16	Q7U242
40	603	30.9	423	16	Q7U114
41	603	30.9	426	16	O05907
42	597.5	30.7	391	16	O05798
43	594.5	30.5	391	16	Q7TX76
44	589	30.2	406	16	P71869
45	589	30.2	406	16	Q7TW99

ALIGNMENTS

RESULT 1

005298	PRELIMINARY;	PRT;	391 AA.
ID	005298		
AC	O05298;		
DT	01-JUL-1997 (TReMBLrel. 04, Created)		
DT	01-JUL-1997 (TReMBLrel. 04, Last sequence update)		
DT	01-OCT-2003 (TReMBLrel. 25, Last annotation update)		
DE	Hypothetical protein (PPE family protein).		
GN	RV1196 OR MTC1364.08 OR MT1234.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37Rv;		
RX	MEDLINE=98295387; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,		
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,		
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne K., Skelton S., Squares S., Rogers J.,		
RA	Rutter S., Seeger K., Taylor K., Whitehead S., Squares R.,		
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence."		
RL	Nature 393:537-544(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CDC 1551 / Oshkosh;		
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,		
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,		
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,		
RA	Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,		
RA	Bisai W.;		
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and		
RT	laboratory strains."		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		

P96362	mycobacteri
Q7U0T5	mycobacteri
Q06386	mycobacteri
Q8VJW0	mycobacteri
P95190	mycobacteri
Q7TX66	mycobacteri
Q7TZ87	mycobacteri
Q7TXX5	mycobacteri
Q3310	mycobacteri
Q33204	mycobacteri
Q8VJZ0	mycobacteri
Q7TZJ5	mycobacteri
Q7TZJ2	mycobacteri
O53940	mycobacteri
O86373	mycobacteri
Q8VJW5	mycobacteri
Q7TX67	mycobacteri
Q9AGF0	mycobacteri
Q7XX3	mycobacteri
Q33312	mycobacteri
Q8VKL9	mycobacteri
Q7U242	mycobacteri
Q7U114	mycobacteri
O05907	mycobacteri
O05798	mycobacteri
Q7TX76	mycobacteri
P71869	mycobacteri
Q7TW99	mycobacteri

DR EMBL: Z93777; CAB07839.1; -.
 DR EMBL: AE007000; AAK45491.1; -.
 DR PIR: B70608; B70608.
 DR TIGR: MT1234; -.
 DR Tuberculist; Rv1196; -.
 DR InterPro; IPR00030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 391 AA; 39158 MW; E409396B8ABDC0F8 CRC64;
 SQ SEQUENCE 391 AA; 39158 MW; E409396B8ABDC0F8 CRC64;
 Query Match 100.0%; Score 1949; DB 16; Length 391;
 Best Local Similarity 100.0%; Pred. No. 2.1e-102;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVSDLSAASAFQSVVWGLTVGWSIG 60
 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVSDLSAASAFQSVVWGLTVGWSIG 60
 Qy 61 SSAGLMVAASPYVAMSVTAQAEELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
 61 SSAGLMVAASPYVAMSVTAQAEELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
 Db 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 Qy 181 LLEQAAVEEASDTAAANQNMNVPAQLQQAQPTQGTTPSSKLGGLWKTVPVPHSPISN 240
 181 LLEQAAVEEASDTAAANQNMNVPAQLQQAQPTQGTTPSSKLGGLWKTVPVPHSPISN 240
 Db 241 MVSANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSGSSG 300
 241 MVSANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSGSSG 300
 Qy 301 LGGGVAANLGRAASVGSLSVQAQWAAANQVTPAARALPLTSLTSAERPGQMLGGLPV 360
 301 LGGGVAANLGRAASVGSLSVQAQWAAANQVTPAARALPLTSLTSAERPGQMLGGLPV 360
 Db 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 Qy 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 Db 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 RESULT 2
 Q7U0E9 PRELIMINARY; PRT; 390 AA.
 ID Q7U0E9
 AC Q7U0E9;
 DT 01-OCT-2003 (TReMBLrel. 25, Created)
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE PPE family protein.
 GN PPE18 OR MB1228.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL: BX248338; CAD94089.1; -.
 KW Complete proteome.
 SQ SEQUENCE 390 AA; 42788276BAB0B436 CRC64;
 Query Match 98.58; Score 1920.5; DB 16; Length 390;
 Best Local Similarity 99.04; Pred. No. 8.4e-101;
 Matches 387; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVSDLSAASAFQSVVWGLTVGWSIG 60
 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVSDLSAASAFQSVVWGLTVGWSIG 60
 Db 61 SSAGLMVAASPYVAMSVTAQAEELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
 61 SSAGLMVAASPYVAMSVTAQAEELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
 Db 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 Qy 181 LLEQAAVEEASDTAAANQNMNVPAQLQQAQPTQGTTPSSKLGGLWKTVPVPHSPISN 240
 181 LLEQAAVEEASDTAAANQNMNVPAQLQQAQPTQGTTPSSKLGGLWKTVPVPHSPISN 240
 Db 241 MVSANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSGSSG 300
 241 MVSANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSGSSG 300
 Qy 301 LGGGVAANLGRAASVGSLSVQAQWAAANQVTPAARALPLTSLTSAERPGQMLGGLPV 360
 301 LGGGVAANLGRAASVGSLSVQAQWAAANQVTPAARALPLTSLTSAERPGQMLGGLPV 360
 Db 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 Qy 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 Db 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 361 GQMGARAGGSLGVLVRPPRYVPHSPAAG 391
 RESULT 3
 Q7U071 PRELIMINARY; PRT; 396 AA.
 ID Q7U071
 AC Q7U071;
 DT 01-OCT-2003 (TReMBLrel. 25, Created)
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE PPE family protein.
 GN PPE19 OR MB1396C.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL: BX248338; CAD94257.1; -.
 KW Complete proteome.
 SQ SEQUENCE 396 AA; 39651 MW; 698A92A50E3CBA3F CRC64;
 Query Match 91.3%; Score 1779.5; DB 16; Length 396;
 Best Local Similarity 91.7%; Pred. No. 7.7e-93;
 Matches 364; Conservative 9; Mismatches 17; Indels 7; Gaps 3;
 Qy 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVSDLSAASAFQSVVWGLTVGWSIG 60
 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVSDLSAASAFQSVVWGLTVGWSIG 60
 Db 61 SSAGLMVAASPYVAMSVTAQAEELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
 61 SSAGLMVAASPYVAMSVTAQAEELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
 Db 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 Qy 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 Db 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 121 LIATNLGONTPTAIIVNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180


```

QY 181 LLEQAAVEEASDTAAANQNMNNVPOALQOAOPTQGTTPSSKLGGLWKTVPSPHRSPI 240
Db 181 LLEQAAVEEASDTAAANQNMNNVPOALQOAOPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMMNNHMTSGVSMNTLSSMLKGFAPAAAAAQVTAQAONGVRAMSS-----LGSSL 296
Db 241 IVSMLNNHVMSTSGVSMNTLSSMLKGFAP-AAAQAVETAQAONGVQAMSSLSGSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPPQAAANQAVTPAARALPLTSLTSAABERGPGOMLG 356
Db 300 GSSGLGGVAAANLGRAASVGSLSVPPQAAANQAVTPAARALPLTSLTSAABERGPGOMLG 359
QY 357 GLPVGQMGARAG--GGLSGVLVPPRPYPMPHSPAAG 391
Db 360 GLPLGH-SVNAGSGINNLRVPRARAYAIPTPTPAAG 393

```

```

RESULT 4
Q8VIZ3 Q8VIZ3 PRELIMINARY; PRT; 393 AA.
AC Q8VIZ3;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE PPE family protein.
GN Mycobacterium tuberculosis.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bhalai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007161; AAK47941.1; -
DR TIGR; MT3582; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE I.1; 1.
SQ SEQUENCE 393 AA; 39658 MW; 96F0B67798855511 CRC64;

```

```

Query Match 84.6%; Score 1648; DB 16; Length 393;
Best Local Similarity 84.8%; Pred. No. 2e-85;
Matches 335; Conservative 20; Mismatches 34; Indels 6; Gaps 3;

QY 1 MVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 1 MVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASAPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
Db 61 SSAGLMVAASAPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
QY 121 LIATNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 180
Db 121 LIATNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNNVPOALQOAOPTQGTTPSSKLGGLWKTVPSPHRSPI 240
Db 181 LLEQAAVEEASDTAAANQNMNNVPOALQOAOPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMMNNHMTSGVSMNTLSSMLKGFAPAAAAAQVTAQAONGVRAMSS-----LGSSL 296
Db 241 MVSMMNNHMTSGVSMNTLSSMLKGFAPAAAAAQVTAQAONGVRAMSS-----LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPPQAAANQAVTPAARALPLTSLTSAABERGPGOMLG 356

```

```

Db 241 IVSMLNNHVMSTSGVSMNTLSSMLKGFAP-AAAQAVETAQAONGVQAMSSLSGSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPPQAAANQAVTPAARALPLTSLTSAABERGPGOMLG 356
Db 300 GSSGLGGVAAANLGRAASVGSLSVPPQAAANQAVTPAARALPLTSLTSAABERGPGOMLG 359
QY 357 GLPVGQMGARAGGGLSGVLVPPRPYPMPHSPAAG 391
Db 360 GLPLGH-SVNAGSGINNLRVPRARAYAIPTPTPAAG 393

RESULT 5
O06341 O06341 PRELIMINARY; PRT; 393 AA.
AC O06341;
DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTCY13E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
DR EMBL; Z95390; CAB08702.1; -
DR PIR; C70568; C70568.
DR Tuberculist; RV3478; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 81.2%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.3e-82;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 1 MVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASAPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
Db 61 SSAGLMVAASAPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
QY 121 LIATNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 180
Db 121 LIATNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNNVPOALQOAOPTQGTTPSSKLGGLWKTVPSPHRSPI 240
Db 181 LLEQAAVEEASDTAAANQNMNNVPOALQOAOPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMMNNHMTSGVSMNTLSSMLKGFAPAAAAAQVTAQAONGVRAMSS-----LGSSL 296
Db 241 MVSMMNNHMTSGVSMNTLSSMLKGFAPAAAAAQVTAQAONGVRAMSS-----LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPPQAAANQAVTPAARALPLTSLTSAABERGPGOMLG 356

```

```
Db 300 GSSGLGAGVAANLGRAASVGSLSVPPAWAANQAVTPPAARALPLTSLTSAQTAPGHMLG 359
QY 357 GLPVQGMGARAGGSLGVLVPPRPVPMHPHSPAAG 391
Db 360 GLPLGH-SVNAAGSGINNLRVPAAYAIPTPPAAG 393

RESULT 6
Q7TWFS PRELIMINARY; PRT; 393 AA.
AC Q7TWFS5
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE PE family protein.
GN PP260 OR MB3505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RA Garnier T., Egliseier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248346; CAD95692.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 81.2%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.3e-82;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDFGALPPEINSARMYAGPSASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIG 60
Db 1 MVDFGALPPEINSARMYAGPSASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLMAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMT 120
QY 121 LIATNLGQNTPAIAVNEAYEGEMWADAAAFYAAATATATATATLPPFEAPMTSAGG 180
QY 121 LTATNLGQNTPAIAVNEAYEGEMWADAAAFYAAATATATATLPPFEAPMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQQAQVPAQQAQVPAQQAQVPAQQAQVPAQQA 240
Db 181 LLEQAAVEEASDTAAANQNMNVPOALQQAQVPAQQAQVPAQQAQVPAQQAQVPAQQA 240
QY 241 MVSMANNHMTNSGVSMNTLSMLKGFAPAPAAAQVQTAQAQVPAQQAQVPAQQAQV 296
Db 241 VSSIANNHMTNSGVSMNTLSMLKGFAPAPAAAQVQTAQAQVPAQQAQVPAQQAQV 299
QY 297 GSSGLGAGVAANLGRAASVGSLSVPPAWAANQAVTPPAARALPLTSLTSAERGPGMLG 356
Db 300 GSSGLGAGVAANLGRAASVGSLSVPPAWAANQAVTPPAARALPLTSLTSAERGPGMLG 359
QY 357 GLPVQGMGARAGGSLGVLVPPRPVPMHPHSPAAG 391
Db 360 GLPLGH-SVNAAGSGINNLRVPAAYAIPTPPAAG 393

RESULT 7
Q7TZJ3 PRELIMINARY; PRT; 393 AA.
AC Q7TZJ3
DT 01-OCT-2003 (TReMBLrel. 25, Created)
```

```
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE PPE family protein.
GN PP26 OR MB1817.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RA Garnier T., Egliseier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 38572 MW; 41944B5E547A8AB0 CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 393;
Best Local Similarity 43.7%; Pred. No. 2e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPSASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGS 61
Db 1 MDVFGALPPEINSARMYAGPSAPVAAASAWNGLAAELSSAATGYETVITQLSSEGWLGP 60
QY 62 SAGLMAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 121
Db 61 ASAAAEAVAPVYVAMNSAAAQAEQAATQARAAAAPFAATVPPPLIAANRASLWQL 120
QY 122 IATNLGQNTPAIAVNEAYEGEMWADAAAFYAAATATATATLPPFEAPMTSAGGL 181
Db 121 ISTNVFGQNTSAIAAAEAQYEGEMWADAAAFYAAATATATLPPFEAPMTSAGGL 179
QY 182 LEOAAVEEASDTAAANQNMNVPOALQQAQVPAQQAQVPAQQAQVPAQQAQVPAQQA 229
Db 180 GTQAAAVATAAGTAAGTATLTEMITGLPNALQSLTSLQSS-NGPLSLWQLFGTNPFT 238
QY 230 -----TVSPHRSPISNMVSMANNHMTNSGVSMNTLSMLKGFAPAPAAAQVQTA 282
Db 239 SISALLTLQPYASPFYNTGLPYFISIGMNNFIOAKTL-GLGSAAPAAVA-----AA 292
QY 283 QNGVRAMSLSGSSLSGSGGVAANLGRAASVGSLSVPOAWAA-ANOAVTPAARALPLT 341
Db 293 GDAAKGLPGLGMLG-----GGPVAAGLGNRAASVGLSVPPVNSGGLPGSVTFGRAPLEVS 348
QY 342 SLTSAERGPQGMGLGVLVQGMGARAGGSLGVLVPPRPVPMHPHSPAAG 391
Db 349 TVSAAPAAAPGSLGGLPL-----AGAGGAGAGP-RYGRPTVMARPPFAG 393

RESULT 8
O53939 PRELIMINARY; PRT; 411 AA.
AC O53939
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-WAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PPE family protein.
GN RV1789 OR MT1838 OR MTV049.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
```

```
RA Gordon S.V., Biglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Rutter S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Sulston J.E., Taylor K., Whitehead S., Squares R.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White C.,
RA Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022021; CAA17711.1; ALT_INIT.
DR EMBL; AB007043; AAK46108.1; -.
DR F1R; G70929; G70929.
DR TIGR; MT1838; -.
DR Tuberculist; RV1789; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 411;
Best Local Similarity 43.7%; Pred. No. 2.1e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSMYAGPGSGLVAAQWDSVADLPSAASAFOSVVMGLTVGWSIGS 61
DB 19 MDFGALPPEINSMYAGPGSGLVAAQWDSVADLPSAASAFOSVVMGLTVGWSIGS 61
QY 62 SAGLMVAASPYVAMSVTAGAQLTAQVRAAAAYETAYGLTVPPVIAENRAELMI 121
DB 79 ASAAAEAVAPYVAMSVTAGAQLTAQVRAAAAYETAYGLTVPPVIAENRAELMI 121
QY 122 IATNLGONTPAIVNAEYGEYEMWQADAAAFYAAATATATATLPPFEAPMTSAGL 181
DB 139 ISTNVFGONTSAIAAAEAQYGEYEMWQADAAAFYAAATATATATLPPFEAPMTSAGL 197
QY 182 LEQAAVEEASDTAAA--NOLMNVPOALQOLAQPTGGTTPSSKLGGLMK----- 229
DB 198 GTQAAVATAGTAQSTLTETITGLPNALQSLTSPQLQSS-NGPLSLWQLFGTFNFFT 256
QY 230 -----TVSPHRSPTSNMVMANNHSMNTSGVMTNLTSSMLKGFAPAAAQAVQTA 282
DB 257 SIGALLTDLPYASFFYNTEGLPYFTSGMGNFIQSAKTL-GLIGSAAPAAVA----- 310
QY 283 QNGVRAMSSIGSSIGSSGLGGVAAANIGRAASVGLSVPOAWAA-ANOAVTPAARALPT 341
DB 311 GDAKGLFGJGGMLG-----GGPVAAGLGNAAVSKLSVPPVWNGPLPGSVTPGAALPVS 366
QY 342 SLTSAAREGPGQMLGGLPVGMQARAGGSLGVLVPRPPRYVMPHSPAAG 391
DB 367 TVSAAPAAFGSLGLGLPL----AGAGGAGAGP-RYGFRTVMARPPFAG 411

RESULT 9
Q99Q11 ID Q99Q11 PRELIMINARY; PRT; 410 AA.
AC Q99Q11;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE RV1808-like protein.
GN MYC1808 OR OV1808.

Gordon S.V., Biglmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Rutter S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Sulston J.E., Taylor K., Whitehead S., Squares R.,
"Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
Nature 393:537-544 (1998).
[2]
SEQUENCE FROM N.A.
STRAIN=CDC 1551 / Oshkosh;
Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White C.,
Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bishai W.;
"Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AL022021; CAA17711.1; ALT_INIT.
EMBL; AB007043; AAK46108.1; -.
F1R; G70929; G70929.
TIGR; MT1838; -.
Tuberculist; RV1789; -.
InterPro; IPR000030; Microbac_PPE.
Pfam; PF00823; PPE; 1.
Complete proteome.
SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 411;
Best Local Similarity 43.7%; Pred. No. 2.1e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSMYAGPGSGLVAAQWDSVADLPSAASAFOSVVMGLTVGWSIGS 61
DB 19 MDFGALPPEINSMYAGPGSGLVAAQWDSVADLPSAASAFOSVVMGLTVGWSIGS 61
QY 62 SAGLMVAASPYVAMSVTAGAQLTAQVRAAAAYETAYGLTVPPVIAENRAELMI 121
DB 79 ASAAAEAVAPYVAMSVTAGAQLTAQVRAAAAYETAYGLTVPPVIAENRAELMI 121
QY 122 IATNLGONTPAIVNAEYGEYEMWQADAAAFYAAATATATATLPPFEAPMTSAGL 181
DB 139 ISTNVFGONTSAIAAAEAQYGEYEMWQADAAAFYAAATATATATLPPFEAPMTSAGL 197
QY 182 LEQAAVEEASDTAAA--NOLMNVPOALQOLAQPTGGTTPSSKLGGLMK----- 229
DB 198 GTQAAVATAGTAQSTLTETITGLPNALQSLTSPQLQSS-NGPLSLWQLFGTFNFFT 256
QY 230 -----TVSPHRSPTSNMVMANNHSMNTSGVMTNLTSSMLKGFAPAAAQAVQTA 282
DB 257 SIGALLTDLPYASFFYNTEGLPYFTSGMGNFIQSAKTL-GLIGSAAPAAVA----- 310
QY 283 QNGVRAMSSIGSSIGSSGLGGVAAANIGRAASVGLSVPOAWAA-ANOAVTPAARALPT 341
DB 311 GDAKGLFGJGGMLG-----GGPVAAGLGNAAVSKLSVPPVWNGPLPGSVTPGAALPVS 366
QY 342 SLTSAAREGPGQMLGGLPVGMQARAGGSLGVLVPRPPRYVMPHSPAAG 391
DB 367 TVSAAPAAFGSLGLGLPL----AGAGGAGAGP-RYGFRTVMARPPFAG 411

RESULT 9
Q99Q11 ID Q99Q11 PRELIMINARY; PRT; 410 AA.
AC Q99Q11;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE RV1808-like protein.
GN MYC1808 OR OV1808.
```

```
OS Mycobacterium microti.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1806;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MYC 94-2272, and OV254;
RA Sivadon V., Heym B., Mazancourt P., Gaillard J.-L.;
RT "PPE RV1808 orthologue of Mycobacterium microti."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF335180; AAK20894.1; -.
DR EMBL; AF335179; AAK20893.1; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
SQ SEQUENCE 410 AA; 40044 MW; 3B9DA3174655A5EA CRC64;

Query Match 38.7%; Score 754; DB 2; Length 410;
Best Local Similarity 44.5%; Pred. No. 5.7e-35;
Matches 177; Conservative 51; Mismatches 140; Indels 30; Gaps 10;

QY 1 MVDFGALPPEINSMYAGPGSGLVAAQWDSVADLPSAASAFOSVVMGLTVGWSIG 60
DB 1 MLDGALPPEINSMYAGPGSGLVAAQWDSVADLPSAASAFOSVVMGLTVGWSIG 60
QY 61 SSAGLMVAASPYVAMSVTAGAQLTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 PSSITMAAAVAPYVAMSVTAGAQLTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTPAIVNAEYGEYEMWQADAAAFYAAATATATATLPPFEAPMTSAG 180
DB 121 LVATNIFGONTPAIVNAEYGEYEMWQADAAAFYAAATATATATLPPFEAPMTSAG 179
QY 181 LLEQAAVEEASDTAAA-----NOLMNVPOALQOLAQPTGGTTPSSKLGGLWK 233
DB 180 TAAQSAVVAAQAAAGAAASDDITAAQ-SQLISLLPSTLQSLA--TTATATSASAG-- 233
QY 234 HRSPTSNMVMANNHSMNTSGV---SMTNTLSSMLKGFAPAAAQAVQTAQNGVRAM 290
DB 234 -LQSTITLANLTGPYSIIIGATPGGHWLFGQIL-GLAQNAGVALLGPKAAAGALS 291
QY 291 SLGSSLG-----SGLGGVAAANIGRAASVGLSVPOAWAAANQAVTPAARALPT 345
DB 292 PLALPLRGYIADITPLGGGATGGIARAIYVGLSVPOGWAEEAAPPVRAVASVLP 351
QY 346 A-AERGPGQMLGGLPVGMQARAGGGL-----SGVLRY 377
DB 352 ALAAEPALFGEMALSSLAGRALLAGTAVRSGAARV 389

RESULT 10
Q93957 ID Q93957 PRELIMINARY; PRT; 409 AA.
AC Q93957;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1808 OR MT1856.1 OR MTV049.30.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Biglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
```


Best Local Similarity 42.8%; Pred. No. 3.2e-34; Matches 182; Conservative 60; Mismatches 122; Indels 61; Gaps 13;

QY 2 VDFGALPPPEINARMYAGPGSASLVAAQWMDVSDLFSAASAFQSVVWGLTVGWSIGS 61
 Db 1 MDAFATLPPPEINARMYAGPGSAPMLAAASAWHGLSALSYSSVLSLTITGEEWHGP 60

QY 62 SAGLVAAASPYVAMSVTAGAELTAQVRVAAAYETAYGLTVPPVPIAENRAELMIL 121
 Db 61 ASASMTAAAPYVAMSVTAVRAEQAGAAQAAAEAAFAATVPPVPIAENRAELMIL 120

QY 122 IATNLLGONTPAIVNAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGL 181
 Db 121 IATNLLGONTPAIVNAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGL 179

QY 182 LEQAAAVEEASDTAAA-----NQLMNVVPOALQQAQPTQGT-----PSSKLG----- 225
 Db 180 AAGSAAIAHATGASAGAQQTTLSQIAAIPSVLQGLSSSTAATASGPGSGLLIGSGSS 239

QY 226 ---GLWKTVPSPHRSPISNMVSMANNHMTNSGVSM-TNTLS---SMLKGFAPAAAQAV 278
 Db 240 WLDKWLALDPN-----SNFWNTIASSGLFLPSNTIAPFLGLLGGVAAADAAGDV 289

QY 279 QTAQNGVRAMSSLSGSSL-----GSSGLGGVAAANLGRAASVGSLSVPOQAAANQAVTP 333
 Db 290 LGATSG-----CLGGALVAPLGSAGLGGTVAGLGNATVGLTSVPSWTAAPLASP 344

QY 334 AARAL-----PLTSITSAARPGQMLGGLPVQMGARAGGGLSVLRVPP-----RPVMPH 386
 Db 345 LGSALGOTPMVAPPVAAAG-----MPGMPFGTMGGQGGF-----RAVPOYGFRRNFVAR 394

QY 387 SPAAG 391
 Db 395 PPAAG 399

RESULT 13

O53956 PRELIMINARY; PRT; 403 AA.

ID O53956
 AC O53956; (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE PPE-family protein.
 GN RV1807 OR MT1856 OR MT049, 29.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Ohkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains";
 Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL022021; CAI17728.1; -;
 DR EMBL; AE007044; AA046128.1; -;
 DR PIR; H70931; H70931.
 DR TIGR; MT1856; -;
 DR TubercuList; RV1807; -;
 DR InterPro; IPR00030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome. 227 F -> S (IN REF. 2).
 FT CONFLICT 238 V -> L (IN REF. 2).
 FT CONFLICT 238
 SQ SEQUENCE 403 AA; 39243 MW; DCE1880FD15CBFE CRC64;

Query Match 37.8%; Score 737.5; DB 16; Length 403;
 Best Local Similarity 42.6%; Pred. No. 4.8e-34;
 Matches 181; Conservative 60; Mismatches 123; Indels 61; Gaps 12;

QY 2 VDFGALPPPEINARMYAGPGSASLVAAQWMDVSDLFSAASAFQSVVWGLTVGWSIGS 61
 Db 5 LDFATLPPPEINARMYAGPGSAPMLAAASAWHGLSALSYSSVLSLTITGEEWHGP 64

QY 62 SAGLVAAASPYVAMSVTAGAELTAQVRVAAAYETAYGLTVPPVPIAENRAELMIL 121
 Db 65 ASASMTAAAPYVAMSVTAVRAEQAGAAQAAAEAAFAATVPPVPIAENRAELMIL 124

QY 122 IATNLLGONTPAIVNAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGL 181
 Db 125 IATNLLGONTPAIVNAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGL 183

QY 182 LEQAAAVEEASDTAAA-----NQLMNVVPOALQQAQPTQGT-----TPSS 222
 Db 184 AAGSAAIAHATGASAGAQQTTLSQIAAIPSVLQGLSSSTAATASGPGSGLLIGSGSS 243

QY 223 KLGGLWKTVPSPHRSPISNMVSMANNHMTNSGVSM-TNTLS---SMLKGFAPAAAQAV 278
 Db 244 WLDKWLALDPN-----SNFWNTIASSGLFLPSNTIAPFLGLLGGVAAADAAGDV 293

QY 279 QTAQNGVRAMSSLSGSSL-----GSSGLGGVAAANLGRAASVGSLSVPOQAAANQAVTP 333
 Db 294 LGATSG-----CLGGALVAPLGSAGLGGTVAGLGNATVGLTSVPSWTAAPLASP 348

QY 334 AARAL-----PLTSITSAARPGQMLGGLPVQMGARAGGGLSVLRVPP-----RPVMPH 386
 Db 349 LGSALGOTPMVAPPVAAAG-----MPGMPFGTMGGQGGF-----RAVPOYGFRRNFVAR 398

QY 387 SPAAG 391
 Db 399 PPAAG 403

RESULT 14

O53950 PRELIMINARY; PRT; 423 AA.

ID O53950
 AC O53950; (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PPE-family protein.
 GN RV1801 OR MT049, 23 OR MT1850.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Ohkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

Sequence 79, Appl
Sequence 80, Appl
Sequence 79, Appl
Sequence 126, App
Sequence 126, App
Sequence 28, Appl
Sequence 341, App
Sequence 341, App
Sequence 333, App
Sequence 333, App
Sequence 321, App
Sequence 321, App
Sequence 329, App
Sequence 329, App
Sequence 325, App
Sequence 325, App
Sequence 196, App
Sequence 196, App

28 983 26.7 355 4 US-09-056-556-79
29 983 26.7 355 4 US-09-072-596-80
30 983 26.7 355 4 US-09-072-967-79
31 766.5 20.8 400 4 US-09-073-009-126
32 766.5 20.8 400 4 US-09-073-010-126
33 743.5 20.2 654 4 US-09-287-849-28
34 743.5 20.2 654 4 US-09-620-412C-341
35 743.5 20.2 654 4 US-09-598-419-341
36 741 20.1 518 4 US-09-620-412C-333
37 741 20.1 518 4 US-09-598-419-333
38 734 19.9 715 4 US-09-620-412C-321
39 734 19.9 715 4 US-09-598-419-321
40 733.5 19.9 715 4 US-09-620-412C-329
41 733.5 19.9 715 4 US-09-598-419-329
42 727 19.7 631 4 US-09-620-412C-325
43 727 19.7 631 4 US-09-598-419-325
44 723.5 19.6 525 4 US-09-556-877-196
45 723.5 19.6 525 4 US-09-620-412C-196

ALIGNMENTS

RESULT 1
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

Query Match 100.0%; Score 3686; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.6e-261;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQSQGGGFAIPICQAWAIAQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGGFAIPICQAWAIAQIRSGGSPVTHIGPTAFGLGVVD 60
QY 61 NNGGARVQVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHFPGDVISVTWQ 120
DB 61 NNGGARVQVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHFPGDVISVTWQ 120
QY 121 TKSQGTGNTVLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAAQWSDSASDL 180
DB 121 TKSQGTGNTVLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAAQWSDSASDL 180
QY 181 FSAASAFOSVVMGLTVGSMWIGSSAGLMVAASPVVAMSVTAGOALTAQVRVAAAAYE 240
DB 181 FSAASAFOSVVMGLTVGSMWIGSSAGLMVAASPVVAMSVTAGOALTAQVRVAAAAYE 240
QY 241 TAYGLTVPVPPVIAENRAELMILIAITNLGQNTPAIVNEAFYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPVPPVIAENRAELMILIAITNLGQNTPAIVNEAFYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEPAPMTSAGGLFQAARVEASTAANOLMNVVPAQLQALQPTCGTT 360
DB 301 ATATATLLPFEPAPMTSAGGLFQAARVEASTAANOLMNVVPAQLQALQPTCGTT 360

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 26.0232 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-16

Perfect score: 3686
Sequence: 1 MHHHHHTAASDNFQSQGG.....SGGPVVMGLGVQVGMNTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/6C_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3686	100.0	729	4	US-09-223-040-2
2	3686	100.0	729	4	US-09-287-849-2
3	2941	79.8	596	4	US-09-287-849-26
4	1967.5	53.4	600	4	US-09-287-849-22
5	1944	52.7	391	3	US-08-818-112-107
6	1944	52.7	391	4	US-08-818-111-102
7	1944	52.7	391	4	US-09-056-556-107
8	1944	52.7	391	4	US-09-072-596-102
9	1944	52.7	391	4	US-09-072-967-107
10	1652.5	44.8	396	3	US-08-818-112-111
11	1652.5	44.8	396	4	US-08-818-111-106
12	1652.5	44.8	396	4	US-09-056-556-111
13	1652.5	44.8	396	4	US-09-072-596-106
14	1652.5	44.8	396	4	US-09-072-967-111
15	1486.5	40.3	359	3	US-08-818-112-109
16	1486.5	40.3	359	4	US-08-818-111-104
17	1486.5	40.3	359	4	US-09-056-556-109
18	1486.5	40.3	359	4	US-09-072-596-104
19	1486.5	40.3	359	4	US-09-072-967-109
20	1184	32.1	263	4	US-09-287-849-8
21	1182	32.1	263	3	US-08-818-112-91
22	1182	32.1	263	4	US-08-818-111-92
23	1182	32.1	263	4	US-09-056-556-91
24	1182	32.1	263	4	US-09-072-596-92
25	1182	32.1	263	4	US-09-072-967-91
26	983	26.7	355	3	US-08-818-112-79
27	983	26.7	355	4	US-08-818-111-80

; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 79.8%; Score 2941; DB 4; Length 596;
Best Local Similarity 99.8%; Pred. No. 7.6e-207;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 202 SSAGLMVAASPYVAMSVTAGQAEILTAQVRAAAAYETAYGLTVPPVPIAENRAELMI 261
Db 69 SSAGLMVAASPYVAMSVTAGQAEILTAQVRAAAAYETAYGLTVPPVPIAENRAELMI 128
QY 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 188
QY 322 LLEQAAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPI 381
Db 189 LLEQAAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPI 248
QY 382 MYSMANNHMTSGVSMNTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
Db 249 MYSMANNHMTSGVSMNTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
QY 442 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 501
Db 309 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
QY 502 GQMGARAGGGLSGVLVPRPYVMPHSPAAAGDIAPPALSQDRFPADPPALPLDPSAMVAQV 561
Db 369 GQMGARAGGGLSGVLVPRPYVMPHSPAAAGDIAPPALSQDRFPADPPALPLDPSAMVAQV 428
QY 562 GPQVWVINTKLGYNNAVAGTGVIDPNGVVLNNHVIAGTDINAFSVGSGQTYGVVDVY 621
Db 429 GPQVWVINTKLGYNNAVAGTGVIDPNGVVLNNHVIAGTDINAFSVGSGQTYGVVDVY 488
QY 622 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPPRAVGRVWALGQT 681
Db 489 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPPRAVGRVWALGQT 548
QY 682 VQASDLSLTGAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGNTAAS 729
Db 549 VQASDLSLTGAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGNTAAS 596

RESULT 4
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US

; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 53.4%; Score 1967.5; DB 4; Length 600;
Best Local Similarity 72.0%; Pred. No. 1e-135;
Matches 431; Conservative 19; Mismatches 82; Indels 67; Gaps 10;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 202 SSAGLMVAASPYVAMSVTAGQAEILTAQVRAAAAYETAYGLTVPPVPIAENRAELMI 261
Db 69 SSAGLMVAASPYVAMSVTAGQAEILTAQVRAAAAYETAYGLTVPPVPIAENRAELMI 128
QY 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 188
QY 322 LLEQAAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPI 381
Db 189 LLEQAAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPI 248
QY 382 MYSMANNHMTSGVSMNTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
Db 249 MYSMANNHMTSGVSMNTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
QY 442 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 501
Db 309 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
QY 502 GQMGARAGGGLSGVLVPRPYVMPHSPAAAGDIAP-----PALSDOR 543
Db 369 GQMGARAGGGLSGVLVPRPYVMPHSPAAAGKLPDVAVINTTCNYGQVWAALNATDFGA 428
QY 544 FADFPALPLDPS-----AMVAQVGPQVNVNNTKLGYNNAVAGTGVIDPN 589
Db 429 AAQFNASPVQSYLRNFLLAAPPPORAAQAOL-QAVPGAAQYIGLVESVAGSCN-----N 482
QY 590 GVVLNNHVIAGTDINAFSVGSGQTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGV 649
Db 483 YELMTINYQFG---DVDA---HGAMIRAQAAASLEAEHCAIVRDVLAAGDFWGA--GSV 533
QY 650 AVGEPVWAMG-----NSGGGGTTPRAVGRVWALGQTVQASDLSLTGAE-ETLNG 697
Db 534 ACQEPITQLGRNFQVYEQANAHGQ-----KVQAAGNNMAQTDSAVGSSWATSNG 583

RESULT 5
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

QY 502 QMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532
Db 361 QMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 8

US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Meto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Wedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-102

Query Match 52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 262 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 321
Db 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPEMTSAGG 180
QY 322 LLEQAAVEEASDTAAANQLMNNVPOALQQAQTOGTTPSSKLGGLWKTVPSPHRSPTISN 381
Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQQAQTOGTTPSSKLGGLWKTVPSPHRSPTISN 240

QY 502 QMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532
Db 361 QMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 7

US-09-056-556-107
; Sequence 107, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND TREATMENT
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-107

Query Match 52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 262 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPEMTSAGG 321
Db 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPEMTSAGG 180
QY 322 LLEQAAVEEASDTAAANQLMNNVPOALQQAQTOGTTPSSKLGGLWKTVPSPHRSPTISN 381
Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQQAQTOGTTPSSKLGGLWKTVPSPHRSPTISN 240
QY 382 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAARQAVOTAAQNGVRAMSSLGSSIGSSG 441
Db 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAARQAVOTAAQNGVRAMSSLGSSIGSSG 300
QY 442 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGPQGMGLGLPV 501
Db 301 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGPQGMGLGLPV 360

QY 382 MVSNNHMTNSGVSMNTLSSMLKGFAPAAQAQVOTAAQNGVRAMSSLSLSSG 441
DB 241 MVSNNHMTNSGVSMNTLSSMLKGFAPAAQAQVOTAAQNGVRAMSSLSLSSG 300
QY 442 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERFGQMLGLPV 501
DB 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERFGQMLGLPV 360
QY 502 GOMGARAGGSLGVLRVPPRYVMPHSPAAG 532
DB 361 GOMGARAGGSLGVLRVPPRYVMPHSPAAG 391

RESULT 9
US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-107

Query Match 52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 142 MVDFGALPPPEINARMYAGPGSASLVAAQMWDSVADLFSAAAFQSVVWGLTVGSWIG 201
DB 1 MVDFGALPPPEINARMYAGPGSASLVAAQMWDSVADLFSAAAFQSVVWGLTVGSWIG 60
QY 202 SSAGLWMAAASPYVWNSVTAQAEITAAQVRVAAAAYETAYGLTVPVIAENRAELMI 261
DB 61 SSAGLWMAAASPYVWNSVTAQAEITAAQVRVAAAAYETAYGLTVPVIAENRAELMI 120

QY 262 LIATNLLQONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLILPFEAPEMTSAGG 321
DB 121 LIATNLLQONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLILPFEAPEMTSAGG 180
QY 322 LLEQAAVEEASDTRAAANQLMNTPQALQOAOPTQGTTPSSKLGGLWKTVSPHRSPTSN 381
DB 181 LLEQAAVEEASDTRAAANQLMNTPQALQOAOPTQGTTPSSKLGGLWKTVSPHRSPTSN 240
QY 382 MVSNNHMTNSGVSMNTLSSMLKGFAPAAQAQVOTAAQNGVRAMSSLSLSSG 441
DB 241 MVSNNHMTNSGVSMNTLSSMLKGFAPAAQAQVOTAAQNGVRAMSSLSLSSG 300
QY 442 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERFGQMLGLPV 501
DB 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERFGQMLGLPV 360
QY 502 GOMGARAGGSLGVLRVPPRYVMPHSPAAG 532
DB 361 GOMGARAGGSLGVLRVPPRYVMPHSPAAG 391

RESULT 10
US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-112-111

Query Match 44.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 142 MVDFGALPPPEINARMYAGPGSASLVAAQMWDSVADLFSAAAFQSVVWGLTVGSWIG 201

Db 1 VDFGALPPEINSAARMYAGPGSASLVAAAKWMDVSADLFSAAATQSVVWGLTTGSGWIG 60
QY 202 SSAGLMVAAASPYYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 61 SSAGLMVAAASPYYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
QY 262 LIATNLLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPFEAPEMTSAGG 321
Db 121 LIATNLLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPLITNPGG 180
QY 322 LLEQAAAVEASDTAAANQLMNNVPQALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPI SN 381
Db 181 LLEQAVAVEAIDTAAANQLMNNVPQALQLOAQPTKSIWPFQDLSLWKAI SPHLSPLSN 240
QY 382 MVSAMNNHMTNSGVSMTNTLSMLKGFAPAAARQAVTAAQNGVRAMSS----LGSSL 437
Db 241 IVSMLNNHVSMTNSGVSMTNTLSMLKGFAPAAA-QAVETAQNGVQAMSSLGSLGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGLSVPOAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGAGVAAANLGRAASVGLSVPOAWAAANQAVTPAARALPLTSLTSAATAPGHMLG 359
QY 498 GLPVQMGARAG--GGLSGVLVPPRPVPMHPSAAG 532
Db 360 GLPLGQLTNSGGGFGVSNALRMPRAYVMPRVPAAG 396
RESULT 11
US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-111-106

Query Match

44.8%; Score 1652.5; DB 4; Length 396;

Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 142 MVDFGALPPEINSAARMYAGPGSASLVAAAKWMDVSADLFSAAATQSVVWGLTVGSGWIG 201
Db 1 VDFGALPPEINSAARMYAGPGSASLVAAAKWMDVSADLFSAAATQSVVWGLTVGSGWIG 60
QY 202 SSAGLMVAAASPYYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 61 SSAGLMVAAASPYYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
QY 262 LIATNLLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 321
Db 121 LIATNLLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPLITNPGG 180
QY 322 LLEQAAAVEASDTAAANQLMNNVPQALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPI SN 381
Db 181 LLEQAVAVEAIDTAAANQLMNNVPQALQLOAQPTKSIWPFQDLSLWKAI SPHLSPLSN 240
QY 382 MVSAMNNHMTNSGVSMTNTLSMLKGFAPAAARQAVTAAQNGVRAMSS----LGSSL 437
Db 241 IVSMLNNHVSMTNSGVSMTNTLSMLKGFAPAAA-QAVETAQNGVQAMSSLGSLGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGLSVPOAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGAGVAAANLGRAASVGLSVPOAWAAANQAVTPAARALPLTSLTSAATAPGHMLG 359
QY 498 GLPVQMGARAG--GGLSGVLVPPRPVPMHPSAAG 532
Db 360 GLPLGQLTNSGGGFGVSNALRMPRAYVMPRVPAAG 396
RESULT 12
US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-111

Query Match 44.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTGVSWIG 201
DB 1 VVDFGALPPEINARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTGVSWIG 60

QY 202 SSAGLMVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
DB 61 SSAGLMVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 262 LIATNLGONTPATVANEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 321
DB 121 LIATNLGONTPATVANEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180

QY 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 381
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240

QY 382 MVSMAHNMSTNGSVMTNTLSMLKGFAPAAAQVQTAQAQNGVRAMS- - - - -LGSSL 437
DB 241 IVSMLNHNVSMTNGSVMASTLHSLMKGFAPAAA-QAVETAQNGVQAMSSLGSLGSSL 299

QY 438 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGQMLG 497
DB 300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGHMLG 359

RESULT 13

US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:

SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-106

Query Match 44.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTGVSWIG 201
DB 1 VVDFGALPPEINARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTGVSWIG 60

QY 202 SSAGLMVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
DB 61 SSAGLMVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 262 LIATNLGONTPATVANEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 321
DB 121 LIATNLGONTPATVANEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180

QY 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 381
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240

QY 382 MVSMAHNMSTNGSVMTNTLSMLKGFAPAAAQVQTAQAQNGVRAMS- - - - -LGSSL 437
DB 241 IVSMLNHNVSMTNGSVMASTLHSLMKGFAPAAA-QAVETAQNGVQAMSSLGSLGSSL 299

QY 438 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGQMLG 497
DB 300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGHMLG 359

RESULT 14

US-09-072-967-111
; Sequence 111, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-967-111

Query Match      44.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 1 VVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
QY 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPETMSAGG 321
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPETMSAGG 180
QY 322 LLEQAAVVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPLSN 381
Db 181 LLEQAVAVEAIDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPLSN 240
QY 382 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAVTAQVTAQNGVRAMS-----LGSSL 437
Db 241 IVSMLNNHVSMTNSGVSMNTLSSMLKGFAPAAA-QAVETAQNGVQAAMSLGSLGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAARPGQWMLG 497
Db 300 GSSGLGAGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAATAPGHMLG 359

RESULT 15
US-08-818-112-109
; Sequence 109, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-818-112-109

Query Match      40.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.3e-101;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 1 VVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMT 120
QY 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPETMSAGG 321
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEAPETMSAGG 180
QY 322 LLEQAAVVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPLSN 381
Db 181 LLEQAVAVEAIDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPLSN 240
QY 382 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAVTAQVTAQNGVRAMS-----LGSSL 437
Db 241 VSSIANNHMSMTNSGVSMNTLSSMLKGFAPAAA-QAVETAQNGVQAAMSLGSLGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAARPGQWMLG 497
Db 300 GSSGLGAGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAATAPGHMLG 359

Search completed: June 22, 2004, 17:27:34
Job time : 28.0232 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 67.5195 Seconds

(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-16

Perfect score: 3686

Sequence: 1 MHHHHTAASDNFOLSOGG.....SCGPFVNLGQVGVNMTAAS 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 28213646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3686	100.0	729	9	US-09-287-849-2
2	3686	100.0	729	12	US-09-886-349A-16
3	3686	100.0	729	14	US-10-359-460-2
4	3686	100.0	729	15	US-10-098-732A-16
5	3686	100.0	729	15	US-10-359-459-2
6	3680	99.8	729	15	US-10-369-983-21
7	3680	99.8	813	15	US-10-369-983-15
8	3680	99.8	825	15	US-10-369-983-14
9	3680	99.8	875	15	US-10-369-983-13
10	3680	99.8	930	14	US-10-098-732A-65
11	3680	99.8	930	15	US-10-369-983-12
12	3680	99.8	1016	15	US-10-369-983-18
13	3680	99.8	1022	15	US-10-369-983-17
14	3680	99.8	1154	15	US-10-369-983-16
15	3677	99.8	729	12	US-09-886-349A-18

16	3677	99.8	729	14	US-10-098-732A-18
17	3677	99.8	729	15	US-10-369-983-22
18	2941	79.8	596	9	US-09-287-849-26
19	2941	79.8	596	12	US-09-886-349A-20
20	2941	79.8	596	14	US-10-359-460-26
21	2941	79.8	596	14	US-10-098-732A-20
22	2637	71.5	1010	15	US-10-369-983-4
23	2631	71.4	723	15	US-10-369-983-2
24	1967.5	53.4	600	9	US-09-287-849-22
25	1967.5	53.4	600	14	US-10-359-460-22
26	1944	52.7	391	12	US-09-886-349A-14
27	1944	52.7	391	14	US-10-193-002-102
28	1944	52.7	391	14	US-10-084-843-107
29	1944	52.7	391	14	US-10-098-732A-14
30	1929	52.3	391	12	US-09-872-186-8
31	1652.5	44.8	396	14	US-10-193-002-106
32	1652.5	44.8	396	14	US-10-084-843-111
33	1583	42.9	393	12	US-10-282-122A-62455
34	1583	42.9	393	12	US-10-282-122A-64892
35	1486.5	40.3	359	14	US-10-193-002-104
36	1486.5	40.3	359	14	US-10-084-843-109
37	1184	32.1	358	9	US-09-287-849-8
38	1184	32.1	358	14	US-10-359-460-8
39	1182	32.1	263	12	US-09-886-349A-12
40	1182	32.1	263	14	US-10-193-002-92
41	1182	32.1	263	14	US-10-084-843-91
42	1182	32.1	263	14	US-10-098-732A-12
43	989	26.8	355	9	US-09-712-363-161
44	987	26.8	330	12	US-09-886-349A-4
45	987	26.8	330	14	US-10-098-732A-4

ALIGNMENTS

RESULT 1

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/819,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 729
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match 100.0%; Score 3686; DB 9; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.4e-244;

Sequence 18, Appl
Sequence 22, Appl
Sequence 26, Appl
Sequence 20, Appl
Sequence 26, Appl
Sequence 20, Appl
Sequence 4, Appl
Sequence 2, Appl
Sequence 22, Appl
Sequence 22, Appl
Sequence 14, Appl
Sequence 102, Appl
Sequence 107, Appl
Sequence 14, Appl
Sequence 8, Appl
Sequence 106, Appl
Sequence 111, Appl
Sequence 62455, A
Sequence 64892, A
Sequence 104, Appl
Sequence 109, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 12, Appl
Sequence 92, Appl
Sequence 91, Appl
Sequence 12, Appl
Sequence 161, Appl
Sequence 4, Appl
Sequence 4, Appl

```

Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFOLSGGGGFAIPGOAIAAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFOLSGGGGFAIPGOAIAAGQIRSGGSPVTHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVRVGSAPASLSGISTGDIVTAVDGPINSATAMADALNGHHPGDIVSVTWQ 120
Db 61 NNGNGARVQVRVGSAPASLSGISTGDIVTAVDGPINSATAMADALNGHHPGDIVSVTWQ 120
QY 121 TKSGGTRTGNVTLAEGPFAEPFMDFGALPPPEINSARMYAGPGSASLVAAQAQWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPFAEPFMDFGALPPPEINSARMYAGPGSASLVAAQAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMILIAITNLGQNTPAIAVNEAIEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMILIAITNLGQNTPAIAVNEAIEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGLLEQAAAVEASDTAAANLMNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGLLEQAAAVEASDTAAANLMNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNVSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNVSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLSGSLGGVAAANLGRASVGSLSVPOAARAAQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSLGGVAAANLGRASVGSLSVPOAARAAQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVQMGARAGGGLSGVLVPPRPYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVQMGARAGGGLSGVLVPPRPYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
QY 601 GATDINAFSVGSGQTYGVVYVDRTQDVAVLQIRGAGGLPSAAIIGGVAVGEPVVMAGN 660
Db 601 GATDINAFSVGSGQTYGVVYVDRTQDVAVLQIRGAGGLPSAAIIGGVAVGEPVVMAGN 660
QY 661 SGGGGTTPRAVGRVWALGQTVQASDSLTGAETLNLGLIOFDAAIQPGDSGGPVVNLGQ 720
Db 661 SGGGGTTPRAVGRVWALGQTVQASDSLTGAETLNLGLIOFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

```

RESULT 2

```

US-09-886-349a-16
; Sequence 16, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-09070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50

```

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 16

; LENGTH: 729

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion

; OTHER INFORMATION: protein MTB72f (Ra12-TB9-Ra35 or MTB32-MTB39

; OTHER INFORMATION: fusion)

US-09-886-349a-16

Query Match 100.0%; Score 3686; DB 12; Length 729;

Best Local Similarity 100.0%; Pred. No. 7.4e-244;

Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPGOAIAAGQIRSGGSPVTHIGPTAFGLGVVD 60

Db 1 MHHHHHTAASDNFOLSGGGGFAIPGOAIAAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVRVGSAPASLSGISTGDIVTAVDGPINSATAMADALNGHHPGDIVSVTWQ 120

Db 61 NNGNGARVQVRVGSAPASLSGISTGDIVTAVDGPINSATAMADALNGHHPGDIVSVTWQ 120

QY 121 TKSGGTRTGNVTLAEGPFAEPFMDFGALPPPEINSARMYAGPGSASLVAAQAQWDSVASDL 180

Db 121 TKSGGTRTGNVTLAEGPFAEPFMDFGALPPPEINSARMYAGPGSASLVAAQAQWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSGWISSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYE 240

Db 181 FSAASAFQSVVWGLTVGSGWISSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYE 240

QY 241 TAYGLTVPPVPIAENRAELMILIAITNLGQNTPAIAVNEAIEYGEWMAQDAAMFGYAAAT 300

Db 241 TAYGLTVPPVPIAENRAELMILIAITNLGQNTPAIAVNEAIEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPEMTSAGLLEQAAAVEASDTAAANLMNVPOALQOLAQPTQGT 360

Db 301 ATATATLLPFEAPEMTSAGLLEQAAAVEASDTAAANLMNVPOALQOLAQPTQGT 360

QY 361 PSSKLGGLWKTVPSPHRSPISNVSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420

Db 361 PSSKLGGLWKTVPSPHRSPISNVSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420

QY 421 TAAQNGVRAMSSLSGSLGGVAAANLGRASVGSLSVPOAARAAQAVTPAARALP 480

Db 421 TAAQNGVRAMSSLSGSLGGVAAANLGRASVGSLSVPOAARAAQAVTPAARALP 480

QY 481 LTSLSAAERPGQMLGGLPVQMGARAGGGLSGVLVPPRPYVMPHSPAAGDIAPPALS 540

Db 481 LTSLSAAERPGQMLGGLPVQMGARAGGGLSGVLVPPRPYVMPHSPAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600

Db 541 QDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600

QY 601 GATDINAFSVGSGQTYGVVYVDRTQDVAVLQIRGAGGLPSAAIIGGVAVGEPVVMAGN 660

Db 601 GATDINAFSVGSGQTYGVVYVDRTQDVAVLQIRGAGGLPSAAIIGGVAVGEPVVMAGN 660

QY 661 SGGGGTTPRAVGRVWALGQTVQASDSLTGAETLNLGLIOFDAAIQPGDSGGPVVNLGQ 720

Db 661 SGGGGTTPRAVGRVWALGQTVQASDSLTGAETLNLGLIOFDAAIQPGDSGGPVVNLGQ 720

QY 721 VVGNTAAS 729

Db 721 VVGNTAAS 729

RESULT 3

```

US-10-359-460-2
; Sequence 2, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:

```

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
PRIOR FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 2
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match 100.0%; Score 3686; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.4e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60

QY 61 NNGNGARQVRVGSAPAAISLGITGDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARQVRVGSAPAAISLGITGDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPVPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPVPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPEMTSAGLLEQAAAEEASDTAAANQLMNNVPQALQQAQPTQGT 360
DB 301 ATATATLLPFEAPEMTSAGLLEQAAAEEASDTAAANQLMNNVPQALQQAQPTQGT 360

QY 361 PSSKLGGLWKTVPVPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 420
DB 361 PSSKLGGLWKTVPVPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 420

QY 421 TAAQNGVRAMSSGLSSGLGGGVAANTGRAASVGSLSVPQAWAANAQAVTPAARALP 480
DB 421 TAAQNGVRAMSSGLSSGLGGGVAANTGRAASVGSLSVPQAWAANAQAVTPAARALP 480

QY 481 LTSLSAERPGQMLGGLPVGQNGARAGGLSGVLVPRPPYVPHSPAGDIAAPPALS 540
DB 481 LTSLSAERPGQMLGGLPVGQNGARAGGLSGVLVPRPPYVPHSPAGDIAAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGFQVNVNITKLYNNNAVAGTGTGIVDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGFQVNVNITKLYNNNAVAGTGTGIVDPNGVLTNNHVA 600

QY 601 GATDINAFSVGSGTGYGVVDVQVAVLQLRGAGGLPSAAIGGVAVGPPVAMGN 660
DB 601 GATDINAFSVGSGTGYGVVDVQVAVLQLRGAGGLPSAAIGGVAVGPPVAMGN 660

QY 661 SGGGGTTPRAVGRVVALGQTVQASDLSLTGABETLNLIOFDAAIQPDGSGPPVNLGQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDLSLTGABETLNLIOFDAAIQPDGSGPPVNLGQ 720

QY 721 VVGNTTAAS 729
DB 721 VVGNTTAAS 729

RESULT 4
US-10-098-732A-16
; Sequence 16, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 16
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein MTB2F (Ra12-1dH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
US-10-098-732A-16

Query Match 100.0%; Score 3686; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.4e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60

QY 61 NNGNGARQVRVGSAPAAISLGITGDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARQVRVGSAPAAISLGITGDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPVPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPVPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPEMTSAGLLEQAAAEEASDTAAANQLMNNVPQALQQAQPTQGT 360
DB 301 ATATATLLPFEAPEMTSAGLLEQAAAEEASDTAAANQLMNNVPQALQQAQPTQGT 360

```

QY 361 PSSKLGGLWKTVPSPHRSPISSNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISSNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTAAARALP 480
DB 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTAAARALP 480
QY 481 LTSLSAERGPQGMGLGGLPVQMGARAGGSLGVLVPPRPYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAERGPQGMGLGGLPVQMGARAGGSLGVLVPPRPYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLYNNVAVGAGTGIVDPNGVVLNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLYNNVAVGAGTGIVDPNGVVLNNHVA 600
QY 601 GATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
DB 601 GATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGMMNTAAS 729
DB 721 VVGMMNTAAS 729

RESULT 5
US-10-359-459-2
; Sequence 2, Application US/10359459
; Publication No. US20040013677A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/10/359,459
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-10-359-459-2

Query Match 100.0%; Score 3686; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.4e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGQFAPIGQAMATAGAIRSGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGQFAPIGQAMATAGAIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAAPASLIGSTGDIITAVDGAPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVVGSAAPASLIGSTGDIITAVDGAPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSGGTGTGNVLAEGPPAEFVMDGALPEINSARWYAGPSASLVAAQWDSVASDL 180
DB 121 TKSGGTGTGNVLAEGPPAEFVMDGALPEINSARWYAGPSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVWGLTYGSGWIGSAGLMVAASPPYVAMSVTAQAEHTAAQVRVAAAAYE 240

```

```

DB 181 FSAASAFQSVWGLTYGSGWIGSAGLMVAASPPYVAMSVTAQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAITNLGQNTPALAVNEAEYEGEMWAOAARAFVAAAT 300
DB 241 TAYGLTVPPVIAENRAELMILIAITNLGQNTPALAVNEAEYEGEMWAOAARAFVAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLEQAAAEEASDSTAANQLMNNVPQALQQAQFTQGT 360
DB 301 ATATATLLPFEAPEMTSAGGLEQAAAEEASDSTAANQLMNNVPQALQQAQFTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISSNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISSNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTAAARALP 480
DB 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTAAARALP 480
QY 481 LTSLSAERGPQGMGLGGLPVQMGARAGGSLGVLVPPRPYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAERGPQGMGLGGLPVQMGARAGGSLGVLVPPRPYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLYNNVAVGAGTGIVDPNGVVLNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLYNNVAVGAGTGIVDPNGVVLNNHVA 600
QY 601 GATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
DB 601 GATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGMMNTAAS 729
DB 721 VVGMMNTAAS 729

RESULT 6
US-10-369-983-21
; Sequence 21, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F
US-10-369-983-21

Query Match 99.8%; Score 3680; DB 15; Length 729;
Best Local Similarity 99.9%; Pred. No. 1.9e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGQFAPIGQAMATAGAIRSGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGQFAPIGQAMATAGAIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAAPASLIGSTGDIITAVDGAPINSATAMADALNGHHPGDVSVTWQ 120

```

Db 61 NNGNGARVQVRVGSAPAAASLGISTGDIITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
QY 121 TKSGETRTGNVTLAEGPPAEFVDFGALPPEINSAKMYAGPGSASLVAAAQWDSVASDL 180
Db 121 TKSGETRTGNVTLAEGPPAEFVDFGALPPEINSAKMYAGPGSASLVAAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGWSIGSAGLMVAASPYVAMMSVTAQAEELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGWSIGSAGLMVAASPYVAMMSVTAQAEELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPVIAENRAELMILIAINLLGQNTPAIAVNEAEYGENWAOQAAAMFYAAAT 300
Db 241 TAYGLTVPPPVIAENRAELMILIAINLLGQNTPAIAVNEAEYGENWAOQAAAMFYAAAT 300
QY 301 ATATATLLPFEAPPEMTSAGGLEQAAAEVEASDTAAANQLMNNVPOALQQLAQTQGT 360
Db 301 ATATATLLPFEAPPEMTSAGGLEQAAAEVEASDTAAANQLMNNVPOALQQLAQTQGT 360
QY 361 PSSKLGGLWKTYSPHRSPISNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTYSPHRSPISNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSGGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSGGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVGQMGARAGGSLGVLVPPPYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVGQMGARAGGSLGVLVPPPYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPDPSSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIIDPENGVLTNHVA 600
Db 541 QDRFADFPALPDPSSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIIDPENGVLTNHVA 600
QY 601 GATDINAFSVGSGQTYGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVGSGQTYGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
Db 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
QY 721 VVGMTAAS 729
Db 721 VVGMTAAS 729

RESULT 7
US-10-369-983-15
; Sequence 15, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DPV)
US-10-369-983-15

Query Match 99.8%; Score 3680; DB 15; Length 813;
Best Local Similarity 99.9%; Pred. No. 2.2e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFOLISQGGGFAIPIGQAWATAGQIRSGGSPYTHIGFTAFILGLGVD 60
Db 1 MHHHHHTAASDNFOLISQGGGFAIPIGQAWATAGQIRSGGSPYTHIGFTAFILGLGVD 60
QY 61 NNGNGARVQVRVGSAPAAASLGISTGDIITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
Db 61 NNGNGARVQVRVGSAPAAASLGISTGDIITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
QY 121 TKSGETRTGNVTLAEGPPAEFVDFGALPPEINSAKMYAGPGSASLVAAAQWDSVASDL 180
Db 121 TKSGETRTGNVTLAEGPPAEFVDFGALPPEINSAKMYAGPGSASLVAAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGWSIGSAGLMVAASPYVAMMSVTAQAEELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGWSIGSAGLMVAASPYVAMMSVTAQAEELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPVIAENRAELMILIAINLLGQNTPAIAVNEAEYGENWAOQAAAMFYAAAT 300
Db 241 TAYGLTVPPPVIAENRAELMILIAINLLGQNTPAIAVNEAEYGENWAOQAAAMFYAAAT 300
QY 301 ATATATLLPFEAPPEMTSAGGLEQAAAEVEASDTAAANQLMNNVPOALQQLAQTQGT 360
Db 301 ATATATLLPFEAPPEMTSAGGLEQAAAEVEASDTAAANQLMNNVPOALQQLAQTQGT 360
QY 361 PSSKLGGLWKTYSPHRSPISNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTYSPHRSPISNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSGGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSGGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVGQMGARAGGSLGVLVPPPYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVGQMGARAGGSLGVLVPPPYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPDPSSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIIDPENGVLTNHVA 600
Db 541 QDRFADFPALPDPSSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIIDPENGVLTNHVA 600
QY 601 GATDINAFSVGSGQTYGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVGSGQTYGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
Db 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
QY 721 VVGMTAAS 729
Db 721 VVGMTAAS 729

RESULT 8
US-10-369-983-14
; Sequence 14, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 825
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-MTI)
US-10-369-983-14

Query Match          99.8%; Score 3680; DB 15; Length 825;
Best Local Similarity 99.9%; Pred. No. 2.3e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSOGGQGFALPIGOAMAIAGQIRSGGSPTVHIGTAFILGLGVD 60
Db 1 MHHHHHTAASDNFQLSOGGQGFALPIGOAMAIAGQIRSGGSPTVHIGTAFILGLGVD 60
QY 61 NNGNGARVQVVGSAAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGNGARVQVVGSAAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TKSOGTGTGNVTLAEGPPAEFMDVFGALPPEINSARMYAGPGSASLVAAQAQMDVSD 180
Db 121 TKSOGTGTGNVTLAEGPPAEFMDVFGALPPEINSARMYAGPGSASLVAAQAQMDVSD 180
QY 181 FSAASAFQSVVWGLTVGSWIGSSAGLMVAASPVVAWMSVTAGQAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSWIGSSAGLMVAASPVVAWMSVTAGQAEITAAQVRAAAAYE 240
QY 241 TAYGLTVPVPIAENRAELMLIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPVPIAENRAELMLIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEEAPEMTSAGLLLEQAAAEASDTPAANQLMNNVPAALQALQPTQGT 360
Db 301 ATATATLLPPEEAPEMTSAGLLLEQAAAEASDTPAANQLMNNVPAALQALQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAERGPQMLGGLPVQMGARAGGLSGVLRVPPRPVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAERGPQMLGGLPVQMGARAGGLSGVLRVPPRPVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVDVDRDQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 600
Db 541 QDRFADFPALPLDPSAMVAQVGVVDVDRDQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 600
QY 601 GATDINAFSVGSGQTVGVVDVDRDQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVGSGQTVGVVDVDRDQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
Db 661 SGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGWNTAAS 729
Db 721 VVGWNTAAS 729
```

RESULT 9

US-10-369-983-13

; Sequence 13, Application US/10369983

; Publication No. US2003023593A1

; GENERAL INFORMATION:

Qy	721	VUGMNTAAS	729
Db	721	VUGMNTAAS	729
RESULT 10			
US-10-098-732A-65			
; Sequence 65, Application US/10098732A			
; Publication No. US20030175294A1			
; GENERAL INFORMATION:			
; APPLICANT: Skeiky, Yasir			
; APPLICANT: Brannon, Mark			
; APPLICANT: Guderian, Jeffrey			
; APPLICANT: Corixa Corporation			
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a			
; FILE REFERENCE: Leishmania Antigen			
; CURRENT FILING DATE: 2003-04-29			
; PRIOR APPLICATION NUMBER: US 60/275,837			
; NUMBER OF SEQ ID NOS: 80			
; SOFTWARE: Patentin Ver. 2.1			
; LENGTH: 930			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence:MTEB72F-MAPS			
; OTHER INFORMATION: (r35f) fusion construct, TB MTEB72F (Ral2-TbH9-Ra35)			
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant			
; OTHER INFORMATION: (TSA or MAPS)			
US-10-098-732A-65			
Query Match 99.8%; Score 3680; DB 14; Length 930;			
Best Local Similarity 99.9%; Pred.No.2.6e-243; Indels 0; Gaps 0;			
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	MHHHHHTAASDNFQLSGGGGFAIPICQAMATAGQIRSGGSPVTHIGTAFILGLGWD	60
Db	1	MHHHHHTAASDNFQLSGGGGFAIPICQAMATAGQIRSGGSPVTHIGTAFILGLGWD	60
Qy	61	NNGNGARVQRVVGSGAPAAASIGISTGDVITAVDGPINISATAMADALNNGHPEDVISVTWQ	120
Db	61	NNGNGARVQRVVGSGAPAAASIGISTGDVITAVDGPINISATAMADALNNGHPEDVISVTWQ	120
Qy	121	TKSGGTTRGNVTLAGPPAEFPWFDFGALPPEINSARMYAGPGSASLVAAAQMWDVSASDL	180
Db	121	TKSGGTTRGNVTLAGPPAEFPWFDFGALPPEINSARMYAGPGSASLVAAAQMWDVSASDL	180
Qy	181	FSAASAFQSVMVGLTVGSSAGLMVAASPYPVAMSVTAGQAELTAQAQVRVAAAAAYE	240
Db	181	FSAASAFQSVMVGLTVGSSAGLMVAASPYPVAMSVTAGQAELTAQAQVRVAAAAAYE	240
Qy	241	TAYGLTVPPPVIABNRaelMTLlATNLlGONTpaIAVNEABYGEMWAQDDAAAMFGYAAT	300
Db	241	TAYGLTVPPPVIABNRaelMTLlATNLlGONTpaIAVNEABYGEMWAQDDAAAMFGYAAT	300
Qy	301	ATAATATLLPFEEAPEMTSAGLLLEQAAAVEASDTAAANQNNNVPOALQLAQTGTT	360
Db	301	ATAATATLLPFEEAPEMTSAGLLLEQAAAVEASDTAAANQNNNVPOALQLAQTGTT	360
Qy	361	PSSKLGLWKTVSPHRSPISNNVSMANNHMSTNSGVSMNTLSMLKGPAAPAARQAVQ	420
Db	361	PSSKLGLWKTVSPHRSPISNNVSMANNHMSTNSGVSMNTLSMLKGPAAPAARQAVQ	420
Qy	421	TAAQNGVRAMSSLGSSSLGGCVGAANLGRAASVGSLSVPQAWAANAQAVTPAAALP	480
Db	421	TAAQNGVRAMSSLGSSSLGGCVGAANLGRAASVGSLSVPQAWAANAQAVTPAAALP	480
Qy	481	LTSLTSAAREPGOMLGGLPYQGOMGARAGGGLSGVLRVPPRPYVNPHSIPAAGDIAPPALS	540
Db	481	LTSLTSAAREPGOMLGGLPYQGOMGARAGGGLSGVLRVPPRPYVNPHSIPAAGDIAPPALS	540

361 PSSKLGGLWKTVPSPHSPISNWVSMANNHMTSGVSMNTLSSMLKGFAPAAAQAVQ 420
421 TAAQNGVRAMSSLGSSSLGSGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALP 480
421 TAAQNGVRAMSSLGSSSLGSGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALP 480
481 LTSLSAERPGQMLGGLPVQMGARAGGGLSGVLRVPPRPYVMPHSPAAAGDIAPPALS 540
481 LTSLSAERPGQMLGGLPVQMGARAGGGLSGVLRVPPRPYVMPHSPAAAGDIAPPALS 540
541 QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
541 QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
661 SGGGGTTPRAVPGRVVAGLQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
661 SGGGGTTPRAVPGRVVAGLQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
721 VVGMMNTAAS 729
721 VVGMMNTAAS 729

RESULT 12

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)

US-10-369-983-18

Query Match 99.8%; Score 3680; DB 15; Length 1016;
Best Local Similarity 99.9%; Pred. No. 2.9e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPQAWAIAQIRSGGSPVHIGPTAFLGLGVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPQAWAIAQIRSGGSPVHIGPTAFLGLGVD 60
QY 61 NNGNGARQVRVGSAPAAASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISTWQ 120
DB 61 NNGNGARQVRVGSAPAAASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TKSGGTTRTGNVTLAEGPPAEFMVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSGGTTRTGNVTLAEGPPAEFMVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSMIGSSAGLWVAASPYVAMSVTAQABLTAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSMIGSSAGLWVAASPYVAMSVTAQABLTAQVRVAAAAYE 240

QY 241 TAYGLTVPVPIAENRAELMLIATNLLGONTPAIVANAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPVPIAENRAELMLIATNLLGONTPAIVANAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEBEAPMTSAGGLEQAAAABEASDTAAANQLMNNVPQALQQAQPTQGT 360
DB 301 ATATATLLPFEBEAPMTSAGGLEQAAAABEASDTAAANQLMNNVPQALQQAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHSPISNWVSMANNHMTSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPSPHSPISNWVSMANNHMTSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSSLGSGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSSLGSGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALP 480
QY 481 LTSLSAERPGQMLGGLPVQMGARAGGGLSGVLRVPPRPYVMPHSPAAAGDIAPPALS 540
DB 481 LTSLSAERPGQMLGGLPVQMGARAGGGLSGVLRVPPRPYVMPHSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
QY 601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
DB 601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRAVPGRVVAGLQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
DB 661 SGGGGTTPRAVPGRVVAGLQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGMMNTAAS 729
DB 721 VVGMMNTAAS 729

RESULT 13

US-10-369-983-17
; Sequence 17, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1022
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB102tm2F (MTB102FTW2, MTB72F-htCC#1)

US-10-369-983-17

Query Match 99.8%; Score 3680; DB 15; Length 1022;
Best Local Similarity 99.9%; Pred. No. 3e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPQAWAIAQIRSGGSPVHIGPTAFLGLGVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPQAWAIAQIRSGGSPVHIGPTAFLGLGVD 60
QY 61 NNGNGARQVRVGSAPAAASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISTWQ 120

Db 61 NNGNGARVQVRVGSAPASLSTGSDVITAVDGPINSATAMADALNGHHPGDVISTVWQ 120
QY 121 TKSGGTGTGNTVLAEGPPAEFVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDL 180
Db 121 TKSGGTGTGNTVLAEGPPAEFVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAASPYVAMSVTAGQAELETAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISAGLMVAASPYVAMSVTAGQAELETAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAEEASDTAAANQLMNNVPQALQQAQTOGTT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAAEEASDTAAANQLMNNVPQALQQAQTOGTT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGGLPVGMQAGARAGGLSGVLVPPRPPYMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGGLPVGMQAGARAGGLSGVLVPPRPPYMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGSQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGN 660
Db 601 GATDINAFSVSGSQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGN 660
QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGSDGPPVNGLGQ 720
Db 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGSDGPPVNGLGQ 720
QY 721 VVGWMTAAS 729
Db 721 VVGWMTAAS 729

RESULT 14
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guiderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009091US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72P-mTCC#2)
US-10-369-983-16

Query Match 99.8%; Score 3680; DB 15; Length 1154;
Best Local Similarity 99.9%; Pred. No. 3.5e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQLSQGGGFAIPGQAWAIAGQIRSGGSPVTHIGTAFGLGVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPGQAWAIAGQIRSGGSPVTHIGTAFGLGVD 60
QY 61 NNGNGARVQVRVGSAPASLSTGSDVITAVDGPINSATAMADALNGHHPGDVISTVWQ 120
Db 61 NNGNGARVQVRVGSAPASLSTGSDVITAVDGPINSATAMADALNGHHPGDVISTVWQ 120
QY 121 TKSGGTGTGNTVLAEGPPAEFVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDL 180
Db 121 TKSGGTGTGNTVLAEGPPAEFVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAASPYVAMSVTAGQAELETAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISAGLMVAASPYVAMSVTAGQAELETAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAEEASDTAAANQLMNNVPQALQQAQTOGTT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAAEEASDTAAANQLMNNVPQALQQAQTOGTT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGGLPVGMQAGARAGGLSGVLVPPRPPYMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGGLPVGMQAGARAGGLSGVLVPPRPPYMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGSQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGN 660
Db 601 GATDINAFSVSGSQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGN 660
QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGSDGPPVNGLGQ 720
Db 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGSDGPPVNGLGQ 720
QY 721 VVGWMTAAS 729
Db 721 VVGWMTAAS 729

RESULT 15
US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20

;; PRIOR APPLICATION NUMBER: US 60/265,737
;; PRIOR FILING DATE: 2001-02-01
;; NUMBER OF SEQ ID NOS: 50
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 18
;; LENGTH: 729
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence.MTB72PMut-SA
;; OTHER INFORMATION: (Ra12-TbHp-Ra35Mut-SA)
US-09-886-349A-18

Query Match 99.8%; Score 3677; DB 12; Length 729;
Best Local Similarity 99.7%; Pred. No. 3.1e-243;
Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSGGGGFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSGGGGFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQRVVGSAPAASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120
DB 61 NNGNGARVQRVVGSAPAASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120

QY 121 TKSGGTRTGNVTLAEGPAPFPWDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDL 180
DB 121 TKSGGTRTGNVTLAEGPAPFPWDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSWIGSSAGLMVAAAAPYVAMWSVTAGQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSWIGSSAGLMVAAAAPYVAMWSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPVPVIAENRAELMILITNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVPVIAENRAELMILITNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPREEAPEMTSAGLLEQAAVEASDTAAANOLMNNVPCALCOLAQTQGT 360
DB 301 ATATATLLPREEAPEMTSAGLLEQAAVEASDTAAANOLMNNVPCALCOLAQTQGT 360

QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420

QY 421 TAAQNGVRAMSSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480

QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGCLSGVLVRPVPYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGCLSGVLVRPVPYVMPHSPAAGDIAPPALS 540

QY 541 QDRFADPPALPLDPSAMVAQVGVVNNINTKLYNNNAVAGTGVITDPNGVLTNNHVIA 600
DB 541 QDRFADPPALPLDPSAMVAQVGVVNNINTKLYNNNAVAGTGVITDPNGVLTNNHVIA 600

QY 601 GATDINAFVSGGQTVGVVDVVDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
DB 601 GATDINAFVSGGQTVGVVDVVDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660

QY 661 SGGGGTPRAVGRVVALGTQVQASDSLGAETLNLGIQFDAAIQPGSGGPFVNVGLQ 720
DB 661 SGGGGTPRAVGRVVALGTQVQASDSLGAETLNLGIQFDAAIQPGSGGPFVNVGLQ 720

QY 721 VVGWNTAAS 729
DB 721 VVGWNTAAS 729

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 67.5195 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-18
Perfect score: 3685
Sequence: 1 MHHHHHTAASDNFOLSGGQ.....AGGPVYVGLGVGMNTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

ALIGNMENTS

RESULT 1

US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009700US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FMTSA
; OTHER INFORMATION: (Ra12-TbHp-Ra3MutsA)
US-09-886-349A-18

Query Match 100.0%; Score 3685; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.4e-239;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFOLSGGQGFALPQQAIAIQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFOLSGGQGFALPQQAIAIQIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGGARVQRVVGSPAASLGISTDVITAVDGFINSATAMADALNHHPCDVTSVTWQ 120

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3685	100.0	729	12	US-09-886-349A-18
2	3685	100.0	729	14	US-10-098-732A-18
3	3685	100.0	729	15	US-10-098-732A-22
4	3682	99.9	729	15	US-10-369-983-21
5	3682	99.9	813	15	US-10-369-983-15
6	3682	99.9	825	15	US-10-369-983-14
7	3682	99.9	875	15	US-10-369-983-13
8	3682	99.9	930	14	US-10-098-732A-65
9	3682	99.9	930	15	US-10-369-983-12
10	3682	99.9	1016	15	US-10-369-983-18
11	3682	99.9	1022	15	US-10-369-983-17
12	3682	99.9	1154	15	US-10-369-983-16
13	3677	99.8	729	9	US-09-287-849-2
14	3677	99.8	729	14	US-09-886-349A-16
15	3677	99.8	729	14	US-10-359-459-2

Db 61 NNGNGARVQRVVGSAPAAASLGISTGCVITAVDGPAINSATAMADALNHHHPGDIIVSTWQ 120
Qy 121 TKSGGTRTGNVTLAGPPPAEFMWDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAGPPPAEFMWDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSSAGLMVAAASPVVAMSVTAGOAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSSAGLMVAAASPVVAMSVTAGOAEITAAQVRAAAAYE 240
Qy 241 TAYGLTVPPPIAENRAELMILATNLGQNTPAIAVNEAEYEGEMWQAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPPIAENRAELMILATNLGQNTPAIAVNEAEYEGEMWQAQDAAMFGYAAAT 300
Qy 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Qy 361 PSSKLGGLWKTIVSPHRSPISNMVSMANNHSMNTNSGVSWTNTLSSMLKGFAPAPAAAQVQ 420
Db 361 PSSKLGGLWKTIVSPHRSPISNMVSMANNHSMNTNSGVSWTNTLSSMLKGFAPAPAAAQVQ 420
Qy 421 TAAQNGVRAMSSLGSSSLGSSGVAANLGRAASVGSLSVPOQAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSSLGSSGVAANLGRAASVGSLSVPOQAAANQAVTPAARALP 480
Qy 481 LTSLSAERPGQMLGGLPVQNGARAGGLSGVLVPRPVPYMPHSPAAAGDIAPPALS 540
Db 481 LTSLSAERPGQMLGGLPVQNGARAGGLSGVLVPRPVPYMPHSPAAAGDIAPPALS 540
Qy 541 QDRFADFPALPLDPSAMVAQVGVQVNVINTKLGYNNAVAGAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVINTKLGYNNAVAGAGTGIVDPNGVLTNNHVA 600
Qy 601 GATDINAFSVGSGQTYGVVDVVDRTODVAVLQIRGAGGLPSAAIGGVAVGEPVVMAGN 660
Db 601 GATDINAFSVGSGQTYGVVDVVDRTODVAVLQIRGAGGLPSAAIGGVAVGEPVVMAGN 660
Qy 661 SGGGGTTPRAVGRVVALGQTVQASDSLTCGAETLNGLIQFDAAIQPDAGGPPVNVGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLTCGAETLNGLIQFDAAIQPDAGGPPVNVGLGQ 720
Qy 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 2

US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US2003017529A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098, 732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275, 837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB/2FmutSA
; OTHER INFORMATION: (Rai2-TbH9-Rai3MutSA)

Query Match 100.0%; Score 3685; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.4e-239;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFOLSQGGGFAIPICQAMAIAGQIRSGGSPTHVHIGTAFILGLVVD 60
Db 1 MHHHHHTAASDNFOLSQGGGFAIPICQAMAIAGQIRSGGSPTHVHIGTAFILGLVVD 60
Qy 61 NNGNGARVQRVVGSAPAAASLGISTGCVITAVDGPAINSATAMADALNHHHPGDIIVSTWQ 120
Db 61 NNGNGARVQRVVGSAPAAASLGISTGCVITAVDGPAINSATAMADALNHHHPGDIIVSTWQ 120
Qy 121 TKSGGTRTGNVTLAGPPPAEFMWDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAGPPPAEFMWDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSSAGLMVAAASPVVAMSVTAGOAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSSAGLMVAAASPVVAMSVTAGOAEITAAQVRAAAAYE 240
Qy 241 TAYGLTVPPPIAENRAELMILATNLGQNTPAIAVNEAEYEGEMWQAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPPIAENRAELMILATNLGQNTPAIAVNEAEYEGEMWQAQDAAMFGYAAAT 300
Qy 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Qy 361 PSSKLGGLWKTIVSPHRSPISNMVSMANNHSMNTNSGVSWTNTLSSMLKGFAPAPAAAQVQ 420
Db 361 PSSKLGGLWKTIVSPHRSPISNMVSMANNHSMNTNSGVSWTNTLSSMLKGFAPAPAAAQVQ 420
Qy 421 TAAQNGVRAMSSLGSSSLGSSGVAANLGRAASVGSLSVPOQAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSSLGSSGVAANLGRAASVGSLSVPOQAAANQAVTPAARALP 480
Qy 481 LTSLSAERPGQMLGGLPVQNGARAGGLSGVLVPRPVPYMPHSPAAAGDIAPPALS 540
Db 481 LTSLSAERPGQMLGGLPVQNGARAGGLSGVLVPRPVPYMPHSPAAAGDIAPPALS 540
Qy 541 QDRFADFPALPLDPSAMVAQVGVQVNVINTKLGYNNAVAGAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVINTKLGYNNAVAGAGTGIVDPNGVLTNNHVA 600
Qy 601 GATDINAFSVGSGQTYGVVDVVDRTODVAVLQIRGAGGLPSAAIGGVAVGEPVVMAGN 660
Db 601 GATDINAFSVGSGQTYGVVDVVDRTODVAVLQIRGAGGLPSAAIGGVAVGEPVVMAGN 660
Qy 661 SGGGGTTPRAVGRVVALGQTVQASDSLTCGAETLNGLIQFDAAIQPDAGGPPVNVGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLTCGAETLNGLIQFDAAIQPDAGGPPVNVGLGQ 720
Qy 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 3

US-10-369-983-22
; Sequence 22, Application US/10369983
; Publication No. US2003023559A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369, 983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357, 351

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB72FmutSA (Mtb72f-mur-SA)
US-10-369-983-22

Query Match      100.0%; Score 3685; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.4e-239;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGQFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGGQFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
QY 61 NNGNGARQVRVGSAPAAASLGISTGVDITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARQVRVGSAPAAASLGISTGVDITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSOGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSOGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVWGLTVGSGWISAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYE 240
DB 181 FSAASAFQSVWGLTVGSGWISAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILITNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVIAENRAELMILITNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEAPEWTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
DB 301 ATATATLLPPEAPEWTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
QY 361 PSSKLGGLWKTVPSPHSPI SNMVSMMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPSPHSPI SNMVSMMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQGMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQGMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIVIDPNGVLTNNHVIA 600
DB 541 QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIVIDPNGVLTNNHVIA 600
QY 601 GATDINAFSVSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660
QY 661 SGGQGGTTPRVPGRVVALGQTVQASDSLTAABETLNGLIQFDDAAIQPDAGGPPVNGLGQ 720
DB 661 SGGQGGTTPRVPGRVVALGQTVQASDSLTAABETLNGLIQFDDAAIQPDAGGPPVNGLGQ 720
QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729

; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F
US-10-369-983-21

Query Match      99.9%; Score 3682; DB 15; Length 729;
Best Local Similarity 99.9%; Pred. No. 5.4e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGQFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGGQFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
QY 61 NNGNGARQVRVGSAPAAASLGISTGVDITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARQVRVGSAPAAASLGISTGVDITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSOGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSOGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVWGLTVGSGWISAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYE 240
DB 181 FSAASAFQSVWGLTVGSGWISAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILITNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVIAENRAELMILITNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEAPEWTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
DB 301 ATATATLLPPEAPEWTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
QY 361 PSSKLGGLWKTVPSPHSPI SNMVSMMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPSPHSPI SNMVSMMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQGMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQGMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIVIDPNGVLTNNHVIA 600
DB 541 QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIVIDPNGVLTNNHVIA 600
QY 601 GATDINAFSVSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660
QY 661 SGGQGGTTPRVPGRVVALGQTVQASDSLTAABETLNGLIQFDDAAIQPDAGGPPVNGLGQ 720
DB 661 SGGQGGTTPRVPGRVVALGQTVQASDSLTAABETLNGLIQFDDAAIQPDAGGPPVNGLGQ 720
QY 721 VVGNTAAS 729
```

RESULT 4

US-10-369-983-21

; Sequence 21, Application US/10369983

; Publication No. US20030235593A1

; GENERAL INFORMATION:

Db 721 VVGMNTAAS 729
|||||
RESULT 5
US-10-369-983-15
; Sequence 15, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DPV)
US-10-369-983-15

Query Match 99.9%; Score 3682; DB 15; Length 813;
Best Local Similarity 99.9%; Pred. No. 6.2e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSGGGFAIPIGQAMAIAGQIRSGGSPTHVIGTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSGGGFAIPIGQAMAIAGQIRSGGSPTHVIGTAFGLGVVD 60
Qy 61 NNGNGARVQVVGSAAPASLGISTGDIVTAVDGAPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGNGARVQVVGSAAPASLGISTGDIVTAVDGAPINSATAMADALNGHHPGDVISTWQ 120
Qy 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSGAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYE 240
Qy 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Qy 301 ATATATLLPPEEAPMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
Db 301 ATATATLLPPEEAPMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
Qy 361 PSSKLGGLWKTVPSPHSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
Qy 421 TAAQNGVRAMSSLGSSLGSGGCGGVAANLGRAASVGSISVPOQAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSGGCGGVAANLGRAASVGSISVPOQAAANQAVTPAARALP 480
Qy 481 LTSLTSAAREGPQMLGGI:PVGMQAGARAGGLSGVLRVPPVPMPSHSPAAGDIAPPALS 540
Db 481 LTSLTSAAREGPQMLGGI:PVGMQAGARAGGLSGVLRVPPVPMPSHSPAAGDIAPPALS 540
Qy 541 QDRFADFPALPDPSAMVAQVGVVNNITKLGYNNAVGAGTGIVDPNGVVLTNHNVIA 600
Db 541 QDRFADFPALPDPSAMVAQVGVVNNITKLGYNNAVGAGTGIVDPNGVVLTNHNVIA 600

Qy 601 GATDINAFSVGSGQTYGYDVWVGVDRTQDVAVLQLRAGAGLPSAAIGGGVAVGCEPVVAMGN 660
Db 601 GATDINAFSVGSGQTYGYDVWVGVDRTQDVAVLQLRAGAGLPSAAIGGGVAVGCEPVVAMGN 660
Qy 661 SGGGGTTRAVPGRVVALGQTVQASDSLSLTGABETLNGLIQPDAAIOPDAGAGPVVNGLQ 720
Db 661 SGGGGTTRAVPGRVVALGQTVQASDSLSLTGABETLNGLIQPDAAIOPDAGAGPVVNGLQ 720
Qy 721 VVGMNTAAS 729
Db 721 VVGMNTAAS 729
|||||

RESULT 6

US-10-369-983-14
; Sequence 14, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 825
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB83F (MTB72F-MTI)
US-10-369-983-14

Query Match 99.9%; Score 3682; DB 15; Length 825;
Best Local Similarity 99.9%; Pred. No. 6.3e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSGGGFAIPIGQAMAIAGQIRSGGSPTHVIGTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSGGGFAIPIGQAMAIAGQIRSGGSPTHVIGTAFGLGVVD 60
Qy 61 NNGNGARVQVVGSAAPASLGISTGDIVTAVDGAPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGNGARVQVVGSAAPASLGISTGDIVTAVDGAPINSATAMADALNGHHPGDVISTWQ 120
Qy 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSGAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYE 240
Qy 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Qy 301 ATATATLLPPEEAPMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
Db 301 ATATATLLPPEEAPMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
Qy 361 PSSKLGGLWKTVPSPHSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
Qy 421 TAAQNGVRAMSSLGSSLGSGGCGGVAANLGRAASVGSISVPOQAAANQAVTPAARALP 480

Db 421 TAAQNGVRAMSLGSLGSLGGVAAANLGRASVGLSVFQAAWAAANQAVTPAARALP 480
 QY 481 LTSLSAAERGGQMLGGLPVQMGARAGGGLSGVLRVPRPYVMPHSPAAAGDIAPFALS 540
 Db 481 LTSLSAAERGGQMLGGLPVQMGARAGGGLSGVLRVPRPYVMPHSPAAAGDIAPFALS 540-
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVIDPVGVLVNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVIDPVGVLVNNHVA 600
 QY 601 GATDINAFSVGSGQTYGVVGVDRDQVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
 Db 601 GATDINAFSVGSGQTYGVVGVDRDQVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
 QY 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
 Db 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
 QY 721 VVGMMNTAAS 729
 Db 721 VVGMMNTAAS 729

RESULT 7
 US-10-369-983-13
 ; Sequence 13, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; PRIOR FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 13
 ; LENGTH: 875
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
 ; OTHER INFORMATION: MTB89F (MTB72F-Erd14)
 US-10-369-983-13

Query Match 99.9%; Score 3682; DB 15; Length 875;
 Best Local Similarity 99.9%; Pred. No. 6.8e-239;
 Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHTAASDNFQSLGSGGQGFAPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 Db 1 MHHHHHTAASDNFQSLGSGGQGFAPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 QY 61 NNGNGARVQVGSAPASLGISTGDIVITAVDGAPINSATAMADALNGHPHGDVSVTWQ 120
 Db 61 NNGNGARVQVGSAPASLGISTGDIVITAVDGAPINSATAMADALNGHPHGDVSVTWQ 120
 QY 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARYAGPGSASLVAAQWMDSVASDL 180
 Db 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARYAGPGSASLVAAQWMDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSMWIGSSAGILWVAASPYVAMVSVTAGQELTAQVEVRAAAVE 240
 Db 181 FSAASAFQSVVWGLTVGSMWIGSSAGILWVAASPYVAMVSVTAGQELTAQVEVRAAAVE 240
 QY 241 TAYGLTVPPVIAENRAELMILIAITNLIGQNTPAIVNEAEYGMWQAQAAAFGVAAT 300
 Db 241 TAYGLTVPPVIAENRAELMILIAITNLIGQNTPAIVNEAEYGMWQAQAAAFGVAAT 300

QY 301 ATATATLLPFEAPEMTSAGLLLEQAAAEASDTAAANQNMNVFQALQOLAQPTGTT 360
 Db 301 ATATATLLPFEAPEMTSAGLLLEQAAAEASDTAAANQNMNVFQALQOLAQPTGTT 360
 QY 361 PSSKLGGLWKTIVSPHRSPISNMVSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAVQ 420
 Db 361 PSSKLGGLWKTIVSPHRSPISNMVSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAVQ 420
 QY 421 TAAQNGVRAMSLGSLGSGGLGGVAAANLGRASVGLSVFQAAWAAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSLGSLGSGGLGGVAAANLGRASVGLSVFQAAWAAANQAVTPAARALP 480
 QY 481 LTSLSAAERGGQMLGGLPVQMGARAGGGLSGVLRVPRPYVMPHSPAAAGDIAPFALS 540
 Db 481 LTSLSAAERGGQMLGGLPVQMGARAGGGLSGVLRVPRPYVMPHSPAAAGDIAPFALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVIDPVGVLVNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVIDPVGVLVNNHVA 600
 QY 601 GATDINAFSVGSGQTYGVVGVDRDQVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
 Db 601 GATDINAFSVGSGQTYGVVGVDRDQVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
 QY 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
 Db 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
 QY 721 VVGMMNTAAS 729
 Db 721 VVGMMNTAAS 729

RESULT 8
 US-10-098-732A-65
 ; Sequence 65, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Brannon, Mark
 ; APPLICANT: Guderian, Jeffrey
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; FILE REFERENCE: 014058-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 65
 ; LENGTH: 930
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
 ; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Rai2-TbH9-Ra35)
 ; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
 ; OTHER INFORMATION: (TSA or MAPS)
 US-10-098-732A-65

Query Match 99.9%; Score 3682; DB 14; Length 930;
 Best Local Similarity 99.9%; Pred. No. 7.3e-239;
 Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHTAASDNFQSLGSGGQGFAPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 Db 1 MHHHHHTAASDNFQSLGSGGQGFAPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 QY 61 NNGNGARVQVGSAPASLGISTGDIVITAVDGAPINSATAMADALNGHPHGDVSVTWQ 120
 Db 61 NNGNGARVQVGSAPASLGISTGDIVITAVDGAPINSATAMADALNGHPHGDVSVTWQ 120

Db 61 NNGNGARVQRVVGSAAPASLIGSTGDIVTAVDGPAINSATAMADALNGHHPGDIVSVTWQ 120
 QY 121 TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
 Db 121 TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMSVTAGQAEELTAQVRAAAAYE 240
 Db 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMSVTAGQAEELTAQVRAAAAYE 240
 QY 241 TAYGLTVPPVIAENRAELMILATNLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPVIAENRAELMILATNLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPPEEAPMTSAGLLEQAAAEASDTAAANQLMNNVPCALQQLAQTQGT 360
 Db 301 ATATATLLPPEEAPMTSAGLLEQAAAEASDTAAANQLMNNVPCALQQLAQTQGT 360
 QY 361 PSSKLGGLWKTVPSPHRSPISNMVSMMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMVSMMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSLSGSGLVAAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSSLSGSSLSGSGLVAAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 QY 481 LTSLSAERPGQMLGGLPVGQMGARAGGLSGVLVPRPPYVMPHSPAAGDIAPPALS 540
 Db 481 LTSLSAERPGQMLGGLPVGQMGARAGGLSGVLVPRPPYVMPHSPAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGVQVNVNINLKGYNNAVAGTGIVDPNGVLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVNINLKGYNNAVAGTGIVDPNGVLTNNHVA 600
 QY 601 GATDINAFSGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGSPVVMGN 660
 Db 601 GATDINAFSGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGSPVVMGN 660
 QY 661 SGGGGTPRAVPRGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGDSGSPVVMGN 720
 Db 661 SGGGGTPRAVPRGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGDSGSPVVMGN 720
 QY 721 VVGNTAAS 729
 Db 721 VVGNTAAS 729

RESULT 9

US-10-369-983-12
 ; Sequence 12, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; SOFTWARE: Patent In Ver. 2.1
 ; NUMBER OF SEQ ID NOS: 22
 ; SEQ ID NO 12
 ; LENGTH: 930
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
 ; OTHER INFORMATION: R95F (MTB72F-MAPS)
 US-10-369-983-12

Query Match 99.9%; Score 3682; DB 15; Length 930;
 Best Local Similarity 99.9%; Pred. No. 7.3e-239;
 Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFILGLVVD 60
 Db 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFILGLVVD 60
 QY 61 NNGNGARVQRVVGSAAPASLIGSTGDIVTAVDGPAINSATAMADALNGHHPGDIVSVTWQ 120
 Db 61 NNGNGARVQRVVGSAAPASLIGSTGDIVTAVDGPAINSATAMADALNGHHPGDIVSVTWQ 120
 QY 121 TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
 Db 121 TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMSVTAGQAEELTAQVRAAAAYE 240
 Db 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMSVTAGQAEELTAQVRAAAAYE 240
 QY 241 TAYGLTVPPVIAENRAELMILATNLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPVIAENRAELMILATNLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPPEEAPMTSAGLLEQAAAEASDTAAANQLMNNVPCALQQLAQTQGT 360
 Db 301 ATATATLLPPEEAPMTSAGLLEQAAAEASDTAAANQLMNNVPCALQQLAQTQGT 360
 QY 361 PSSKLGGLWKTVPSPHRSPISNMVSMMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMVSMMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSLSGSGLVAAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSSLSGSSLSGSGLVAAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 QY 481 LTSLSAERPGQMLGGLPVGQMGARAGGLSGVLVPRPPYVMPHSPAAGDIAPPALS 540
 Db 481 LTSLSAERPGQMLGGLPVGQMGARAGGLSGVLVPRPPYVMPHSPAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGVQVNVNINLKGYNNAVAGTGIVDPNGVLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVNINLKGYNNAVAGTGIVDPNGVLTNNHVA 600
 QY 601 GATDINAFSGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGSPVVMGN 660
 Db 601 GATDINAFSGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGSPVVMGN 660
 QY 661 SGGGGTPRAVPRGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGDSGSPVVMGN 720
 Db 661 SGGGGTPRAVPRGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGDSGSPVVMGN 720
 QY 721 VVGNTAAS 729
 Db 721 VVGNTAAS 729

RESULT 10

US-10-369-983-18
 ; Sequence 18, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Gudarian, Jeff
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22


```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match          99.9%; Score 3682; DB 15; Length 1016;
Best Local Similarity 99.9%; Pred. No. 8.2e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSGGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVD 60
DB 1 MHHHHHTAASDNFQLSGGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVD 60
QY 61 NNGNGARVQVRVGSAPAAASLGISTGSDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVRVGSAPAAASLGISTGSDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TSGGTRTGNVTLAEGPPAEFVDFGALPPEINARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TSGGTRTGNVTLAEGPPAEFVDFGALPPEINARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGSAGLVAAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGSAGLVAAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPPVIAENRAELMILIAIATNLGQNTPAIAYNEAEYGEWMAQDAAMFYAAAT 300
DB 241 TAYGLTVPPPPVIAENRAELMILIAIATNLGQNTPAIAYNEAEYGEWMAQDAAMFYAAAT 300
QY 301 ATATATLLPPEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQQAQTOGTT 360
DB 301 ATATATLLPPEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQQAQTOGTT 360
QY 361 PSSKLGGLWKTVSPHRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVSPHRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
QY 481 LTSLTSAERPGQMLGGLPVGMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLTSAERPGQMLGGLPVGMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTTNNHVA 600
QY 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAALGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAALGGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQFGDAGGPPVNGLQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQFGDAGGPPVNGLQ 720
QY 721 VVGMNTAAS 729
DB 721 VVGMNTAAS 729

RESULT 11
US-10-369-983-17
; Sequence 17, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir

```

```

; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1022
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB102tm2f (MTB102fTM2, MTB72F-HTCC#1)
US-10-369-983-17

Query Match          99.9%; Score 3682; DB 15; Length 1022;
Best Local Similarity 99.9%; Pred. No. 8.3e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSGGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVD 60
DB 1 MHHHHHTAASDNFQLSGGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVD 60
QY 61 NNGNGARVQVRVGSAPAAASLGISTGSDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVRVGSAPAAASLGISTGSDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TSGGTRTGNVTLAEGPPAEFVDFGALPPEINARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TSGGTRTGNVTLAEGPPAEFVDFGALPPEINARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGSAGLVAAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGSAGLVAAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPPVIAENRAELMILIAIATNLGQNTPAIAYNEAEYGEWMAQDAAMFYAAAT 300
DB 241 TAYGLTVPPPPVIAENRAELMILIAIATNLGQNTPAIAYNEAEYGEWMAQDAAMFYAAAT 300
QY 301 ATATATLLPPEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQQAQTOGTT 360
DB 301 ATATATLLPPEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQQAQTOGTT 360
QY 361 PSSKLGGLWKTVSPHRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVSPHRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLSGSLGSGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLSGSLGSGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
QY 481 LTSLTSAERPGQMLGGLPVGMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLTSAERPGQMLGGLPVGMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTTNNHVA 600
QY 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAALGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAALGGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQFGDAGGPPVNGLQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQFGDAGGPPVNGLQ 720
QY 721 VVGMNTAAS 729

```

Db	721	VVGMNTAAS	729		660
RESULT 12					660
US-10-369-983-16					660
; Sequence 16, Application US/10369983					660
; Publication No. US20030235593A1					660
; GENERAL INFORMATION:					660
; APPLICANT: Skeiky, Yasir					660
; APPLICANT: Guderian, Jeff					660
; APPLICANT: Reed, Steven					660
; APPLICANT: Corixa Corporation					660
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis					660
; FILE REFERENCE: 014058-009081US					660
; CURRENT APPLICATION NUMBER: US/10/369,983					660
; CURRENT FILING DATE: 2003-02-18					660
; PRIOR APPLICATION NUMBER: US 60/357,351					660
; PRIOR FILING DATE: 2002-02-15					660
; NUMBER OF SEQ ID NOS: 22					660
; SOFTWARE: PatentIn Ver. 2.1					660
; SEQ ID NO 16					660
; LENGTH: 1154					660
; TYPE: PRT					660
; ORGANISM: Artificial Sequence					660
; FEATURE:					660
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein					660
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)					660
US-10-369-983-16					660
Query Match 99.9%; Score 3682; DB 15; Length 1154;					660
Best Local Similarity 99.9%; Pred. No. 9,7e-239;					660
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;					660
QY	1	MHHHHHTAASDNFOLSQGGGFAIPICQMAIAGQIRSGGSPVTHIGPTAFGLGVVD	60		660
Db	1	MHHHHHTAASDNFOLSQGGGFAIPICQMAIAGQIRSGGSPVTHIGPTAFGLGVVD	60		660
QY	61	NNNGGARVORVVGSAASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ	120		660
Db	61	NNNGGARVORVVGSAASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ	120		660
QY	121	TKSGGTRTGNVTLAEGPPAEFMVDFGALPPEINSARMYAGPGSASLVAAQMDVSASDL	180		660
Db	121	TKSGGTRTGNVTLAEGPPAEFMVDFGALPPEINSARMYAGPGSASLVAAQMDVSASDL	180		660
QY	181	FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAYE	240		660
Db	181	FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAYE	240		660
QY	241	TAYGLTVPPVPIAENRAELMILITNLGQNTPAIAVNEAEYEGMWAQDAAMFGYAAAT	300		660
Db	241	TAYGLTVPPVPIAENRAELMILITNLGQNTPAIAVNEAEYEGMWAQDAAMFGYAAAT	300		660
QY	301	ATATATLTPPEEAPMTSAGGLEQAAVEASDTAAANQIMNNVPAQLQQLAQPTGGTT	360		660
Db	301	ATATATLTPPEEAPMTSAGGLEQAAVEASDTAAANQIMNNVPAQLQQLAQPTGGTT	360		660
QY	361	PSSKLGGLWKTVPHSRSPISNNVMANNHMTNSGVSMNTLTSSMLKGPAPAAAQAVQ	420		660
Db	361	PSSKLGGLWKTVPHSRSPISNNVMANNHMTNSGVSMNTLTSSMLKGPAPAAAQAVQ	420		660
QY	421	TAAQNGVRAMSLGSSLSGGLGGVVAANIGRAASVGSLSVPQAWAANAQVTPAARALP	480		660
Db	421	TAAQNGVRAMSLGSSLSGGLGGVVAANIGRAASVGSLSVPQAWAANAQVTPAARALP	480		660
QY	481	LTSLTSAAEERPGQMLGGLPVQMGARAGGGLSGVLVPRPPYVPHSPAAAGIAPPAALS	540		660
Db	481	LTSLTSAAEERPGQMLGGLPVQMGARAGGGLSGVLVPRPPYVPHSPAAAGIAPPAALS	540		660
QY	541	QDRFADFALPLDPSAMVAQVGPQVNTIKLGYNNVAVAGTGTVIDPNGVITNNNEVIA	600		660
Db	541	QDRFADFALPLDPSAMVAQVGPQVNTIKLGYNNVAVAGTGTVIDPNGVITNNNEVIA	600		660

QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQLOAQTQTT 360
 DB 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQLOAQTQTT 360
 QY 361 PSSKLGGLWKTVPSPHRSPTSNMVMANNHMTNSGVSTNTLSSMLKGFAPAAAQAVQ 420
 DB 361 PSSKLGGLWKTVPSPHRSPTSNMVMANNHMTNSGVSTNTLSSMLKGFAPAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLGSSSLGSSGLGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLGSSSLGSSGLGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVQOMGARAGGSLGSLRVPPRYMPPHSPAAGDIAPPALS 540
 DB 481 LTSLSAAERGPQMLGGLPVQOMGARAGGSLGSLRVPPRYMPPHSPAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAGAGTGIVDPNGVVLNNHVA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAGAGTGIVDPNGVVLNNHVA 600
 QY 601 GATDINAFSVSGQTYGVDPVQVVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 660
 DB 601 GATDINAFSVSGQTYGVDPVQVVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 660
 QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQ 720
 DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQ 720
 QY 721 VVGMMNTAAS 729
 DB 721 VVGMMNTAAS 729

RESULT 14
 US-09-886-349A-16
 ; Sequence 16, Application US/09886349A
 ; Publication No. US20040086523A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009070US
 ; CURRENT APPLICATION NUMBER: US/09/886,349A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: US 09/597,796
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: US 60/265,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 16
 ; LENGTH: 729
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: tti-fusion
 ; OTHER INFORMATION: Protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
 ; OTHER INFORMATION: Fusion)
 ; US-09-886-349A-16

Query Match 99.8%; Score 3677; DB 12; Length 729;
 Best Local Similarity 99.7%; Pred. No. 1.2e-238;
 Matches 72; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDFLSQSGGQFAIPGQAMATAGQIRSGGSPVTHIGPTAFLGLGVD 60
 DB 1 MHHHHHTAASDFLSQSGGQFAIPGQAMATAGQIRSGGSPVTHIGPTAFLGLGVD 60
 QY 61 NNGNGARVQVRVGSAPAAASLGISTGDIVTAVDQAPINSATAMADALNHHGHPGDIVSVTWQ 120

DB 61 NNGNGARVQVRVGSAPAAASLGISTGDIVTAVDQAPINSATAMADALNHHGHPGDIVSVTWQ 120
 QY 121 TKSGGTRTGNVTLLAEGPPAEFMVDFGALPPEINSAEWYAGPGSASLVAAAQWDSVASDL 180
 DB 121 TKSGGTRTGNVTLLAEGPPAEFMVDFGALPPEINSAEWYAGPGSASLVAAAQWDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASAPYVAMSVTAQAEELTAQVRAAAAYE 240
 DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASAPYVAMSVTAQAEELTAQVRAAAAYE 240
 QY 241 TAYGLTVPPVPIAENRAELMILITATNLLGQNTTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 DB 241 TAYGLTVPPVPIAENRAELMILITATNLLGQNTTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQLOAQTQTT 360
 DB 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNNVPOALQLOAQTQTT 360
 QY 361 PSSKLGGLWKTVPSPHRSPTSNMVMANNHMTNSGVSTNTLSSMLKGFAPAAAQAVQ 420
 DB 361 PSSKLGGLWKTVPSPHRSPTSNMVMANNHMTNSGVSTNTLSSMLKGFAPAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLGSSSLGSSGLGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLGSSSLGSSGLGGVVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVQOMGARAGGSLGSLRVPPRYMPPHSPAAGDIAPPALS 540
 DB 481 LTSLSAAERGPQMLGGLPVQOMGARAGGSLGSLRVPPRYMPPHSPAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAGAGTGIVDPNGVVLNNHVA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAGAGTGIVDPNGVVLNNHVA 600
 QY 601 GATDINAFSVSGQTYGVDPVQVVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 660
 DB 601 GATDINAFSVSGQTYGVDPVQVVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 660
 QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQ 720
 DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQ 720
 QY 721 VVGMMNTAAS 729
 DB 721 VVGMMNTAAS 729

RESULT 15
 US-10-359-460-2
 ; Sequence 2, Application US/10359460
 ; Publication No. US20030147911A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; TITLE OF INVENTION: and Their Uses
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/10/359,460
 ; CURRENT FILING DATE: 2003-02-05
 ; PRIOR APPLICATION NUMBER: US/09/287,849
 ; PRIOR FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07

; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match 99.8%; Score 3677; DB 14; Length 729;
Best Local Similarity 99.7%; Pred. No. 1.2e-238;
Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MHHHHHTAASDNFQLSGGGGFAIPIGQAWAIAQIRSGGSGPTVHIGTAFGLGVVD	60
Db	1	MHHHHHTAASDNFQLSGGGGFAIPIGQAWAIAQIRSGGSGPTVHIGTAFGLGVVD	60
Qy	61	NNNGARVQRVVGSAPASLGISTGDIVITAVDGAIPINSATAMADALNCHHPGDVISVTWQ	120
Db	61	NNNGARVQRVVGSAPASLGISTGDIVITAVDGAIPINSATAMADALNCHHPGDVISVTWQ	120
Qy	121	TKSGGTRTGNVTLAEGPPAFPMVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL	180
Db	121	TKSGGTRTGNVTLAEGPPAFPMVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL	180
Qy	181	FSAASAFQSVVWGLTVGSGWISGAGLVAAASPVVAMSVTAGQAEITAAQVEVAAAAYE	240
Db	181	FSAASAFQSVVWGLTVGSGWISGAGLVAAASPVVAMSVTAGQAEITAAQVEVAAAAYE	240
Qy	241	TAYGLTVPPVPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT	300
Db	241	TAYGLTVPPVPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT	300
Qy	301	ATATATLLPFEAPEMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQAQPTQGT	360
Db	301	ATATATLLPFEAPEMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQAQPTQGT	360
Qy	361	PSSKLGGLWKTSPHRSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQ	420
Db	361	PSSKLGGLWKTSPHRSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQ	420
Qy	421	TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAANAQAVTPAARALP	480
Db	421	TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAANAQAVTPAARALP	480
Qy	481	LTSITSAERGPQMLGGLPVQGMGARAGGGLSGVLVPPRPVPMPSAAGDIAPPALS	540
Db	481	LTSITSAERGPQMLGGLPVQGMGARAGGGLSGVLVPPRPVPMPSAAGDIAPPALS	540
Qy	541	QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIVIDPNGVLTNNHVA	600
Db	541	QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIVIDPNGVLTNNHVA	600
Qy	601	GATDINAFSVGSGQTGVGVVVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN	660
Db	601	GATDINAFSVGSGQTGVGVVVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN	660
Qy	661	SGCGGTTPRAVGRVVALQGTQVADSLSLTCAETLNGLIQFDAAIQPDGAGGVVNGLGQ	720
Db	661	SGCGGTTPRAVGRVVALQGTQVADSLSLTCAETLNGLIQFDAAIQPDGAGGVVNGLGQ	720
Qy	721	VVGWNTAAS 729	
Db	721	VVGWNTAAS 729	

Sequence 79, Appl
Sequence 80, Appl
Sequence 79, Appl
Sequence 126, App
Sequence 126, App
Sequence 16, Appl
Sequence 12, Appl
Sequence 142, App
Sequence 142, App
Sequence 131, App
Sequence 15, Appl
Sequence 57, Appl
Sequence 208, App
Sequence 92, Appl
Sequence 114, App
Sequence 109, App
Sequence 114, App

28 983 32.7 355 4 US-09-056-556-79
29 983 32.7 355 4 US-09-072-596-80
30 983 32.7 355 4 US-09-072-967-79
31 770.5 25.6 400 4 US-09-073-009-126
32 770.5 25.6 400 4 US-09-073-010-126
33 607.5 20.2 710 4 US-09-287-849-16
34 607.5 20.2 856 4 US-09-287-849-12
35 606.5 20.2 423 4 US-09-073-009-142
36 606.5 20.2 423 4 US-09-073-010-142
37 450.5 15.0 943 4 US-09-477-135A-131
38 424 14.1 141 4 US-09-073-009-15
39 424 14.1 141 4 US-09-073-010-15
40 381.5 12.7 204 4 US-08-311-731A-57
41 377.5 12.6 208 4 US-08-311-731A-208
42 328.5 10.9 371 4 US-09-050-739-92
43 325.5 10.8 368 3 US-08-818-112-114
44 325.5 10.8 368 4 US-08-818-111-109
45 325.5 10.8 368 4 US-09-056-556-114

ALIGNMENTS

RESULT 1
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849.
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 100.0%; Score 3007; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.3e-219;
Matches 596; Conservative 0; Mismatches 0; Gaps 0;

QY 1 HMHHHHHMDVFGALPPEINSARMYAGPGSASLVAAQMDVSASDLFSAASAFQSVVWG 60
Db 1 HMHHHHHMDVFGALPPEINSARMYAGPGSASLVAAQMDVSASDLFSAASAFQSVVWG 60

QY 61 LTVGSNISSAGLVAAASPVVANSVTAGAEITAAQVRVAAAYETAYGLTVPFPVIA 120
Db 61 LTVGSNISSAGLVAAASPVVANSVTAGAEITAAQVRVAAAYETAYGLTVPFPVIA 120

QY 121 ENRAELMILITATNLIGONTPTAIVNNEAEYGENWQAQDAAMFGYAAATATATLPPFEA 180

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 21.2754 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-20
Perfect score: 3007
Sequence: 1 HMHHHHHMDVFGALPPEIN.....SGGPVNGLGQVGMNTAAS 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/aaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/aaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/aaa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/aaa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/aaa/PCUS_COMB.pep:*
6: /cgn2_6/prodata/2/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3007	100.0	596	4	US-09-287-849-26
2	2941	97.8	729	4	US-09-223-040-2
3	2941	97.8	729	4	US-09-287-849-2
4	2033.5	67.6	600	4	US-09-287-849-22
5	1949	64.8	391	3	US-08-818-112-107
6	1949	64.8	391	4	US-08-818-111-102
7	1949	64.8	391	4	US-09-056-556-107
8	1949	64.8	391	4	US-09-072-596-102
9	1949	64.8	391	4	US-09-072-596-107
10	1652.5	55.0	396	3	US-08-818-111-111
11	1652.5	55.0	396	4	US-08-818-111-106
12	1652.5	55.0	396	4	US-09-056-556-111
13	1652.5	55.0	396	4	US-09-072-596-106
14	1652.5	55.0	396	4	US-09-072-967-111
15	1486.5	49.4	359	3	US-08-818-111-109
16	1486.5	49.4	359	4	US-08-818-111-104
17	1486.5	49.4	359	4	US-09-056-556-109
18	1486.5	49.4	359	4	US-09-056-556-104
19	1486.5	49.4	359	4	US-08-072-967-109
20	1189	39.5	388	4	US-09-287-849-8
21	1187	39.5	263	3	US-08-818-112-91
22	1187	39.5	263	4	US-08-818-111-92
23	1187	39.5	263	4	US-09-056-556-91
24	1187	39.5	263	4	US-09-072-596-92
25	1187	39.5	263	4	US-09-072-967-91
26	983	32.7	355	3	US-08-818-112-79
27	983	32.7	355	4	US-08-818-111-60

Db 121 ENRAELMILIAITNLGQNTPAIAVNEAYEGMWAQDAAMFGYAAATATATATALLPFEA 180
QY 181 PEMTSAGLLBQAAAVEASDTAAANQLMNNVPALQLOAQPTQGTTPSSKLGGLWKTVS 240
Db 181 PEMTSAGLLBQAAAVEASDTAAANQLMNNVPALQLOAQPTQGTTPSSKLGGLWKTVS 240
QY 241 PHRSPISNMWMANHSMNTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSL 300
Db 241 PHRSPISNMWMANHSMNTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSL 300
QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAASRGPG 360
Db 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAASRGPG 360
QY 361 QMLGGLPVGQMGARAGGGLSVLRVPPRYMPHSPAAGDIAPPALSQDRFADFPALPLD 420
Db 361 QMLGGLPVGQMGARAGGGLSVLRVPPRYMPHSPAAGDIAPPALSQDRFADFPALPLD 420
QY 421 PSAMVAQVGPQVNNINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
Db 421 PSAMVAQVGPQVNNINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
QY 481 QTYGVDVYDRTQDQAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
Db 481 QTYGVDVYDRTQDQAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
QY 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
Db 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596

RESULT 2
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-223-040-2

Query Match 97.8%; Score 2941; DB 4; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-214;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQMDVSDVLSAASAFQSVVWGLTVGSMIG 68
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAQMDVSDVLSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLMVAASPYVAMSVTAQAEITAAQVRAAAAYETAYGLTVPPVPIAENRAELMI 128
Db 202 SSAGLMVAASPYVAMSVTAQAEITAAQVRAAAAYETAYGLTVPPVPIAENRAELMI 261

QY 129 LIATNLGQNTPAIAVNEAYEGMWAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 188
Db 262 LIATNLGQNTPAIAVNEAYEGMWAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 321

QY 189 LLEQAAAVEASDTAAANQLMNNVPALQLOAQPTQGTTPSSKLGGLWKTVSPIHRSPISN 248

Db 322 LLEQAAAVEASDTAAANQLMNNVPALQLOAQPTQGTTPSSKLGGLWKTVSPIHRSPISN 381
QY 249 MYSMANNHSMNTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
Db 382 MYSMANNHSMNTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
QY 309 LGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAASRGPGQMLGGLPV 368
Db 442 LGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAASRGPGQMLGGLPV 501
QY 369 GQMGARAGGGLSVLRVPPRYMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGGLSVLRVPPRYMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNNINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDV 488
Db 562 GPQVNNINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDV 621
QY 489 GYDRTQDQAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPGVVALGQT 548
Db 622 GYDRTQDQAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPGVVALGQT 681
QY 549 VOASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
Db 682 VOASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 729

RESULT 3
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match 97.8%; Score 2941; DB 4; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-214;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQMDVSDVLSAASAFQSVVWGLTVGSMIG 68
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAQMDVSDVLSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLMVAASPYVAMSVTAQAEITAAQVRAAAAYETAYGLTVPPVPIAENRAELMI 128

Db 202 SSAGLWAAAGPYVAMSVTAQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
Qy 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
Db 262 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 321
Qy 189 LLEQAAAEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYSPHRSPISN 248
Db 322 LLEQAAAEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYSPHRSPISN 381
Qy 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAQVCTAAQNGVRAMSSLSGSSG 308
Db 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAQVCTAAQNGVRAMSSLSGSSG 441
Qy 309 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAARERGQMLGGLPV 368
Db 442 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAARERGQMLGGLPV 501
Qy 369 GQMGARAGGGLSGVLRRPVPYMPHSPAAGDIAPPALSQRRPAPFPALPDPSAMVAQV 428
Db 502 GQMGARAGGGLSGVLRRPVPYMPHSPAAGDIAPPALSQRRPAPFPALPDPSAMVAQV 561
Qy 429 GPQVNNITKLYNNNAVGAGTGIVIDPENGVLTNHVIAGATDINAFSVGSGQTYGVDDV 488
Db 562 GPQVNNITKLYNNNAVGAGTGIVIDPENGVLTNHVIAGATDINAFSVGSGQTYGVDDV 621
Qy 489 GYDRTQDVAVLQIRGAGGLPSAAIGGVAVGVPVAMGNSGGQGTTPRAVGRVVALQOT 548
Db 622 GYDRTQDVAVLQIRGAGGLPSAAIGGVAVGVPVAMGNSGGQGTTPRAVGRVVALQOT 681
Qy 549 VQASDLSLTGABETLNGLIQFDAAIOPGDSGGVPVNGLGQVGMNTAAS 596
Db 682 VQASDLSLTGABETLNGLIQFDAAIOPGDSGGVPVNGLGQVGMNTAAS 729

RESULT 4
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 67.6%; Score 2033.5; DB 4; Length 600;
Best Local Similarity 72.5%; Pred. No. 6.7e-146;

Matches 440; Conservative 19; Mismatches 81; Indels 67; Gaps 10;
Qy 1 HMEHHHHHVVDFGALPEINSARMYAGPGSASILVAAAQWMDSVASDLFSAASAFQSVVWG 60
Db 1 HMEHHHHHVVDFGALPEINSARMYAGPGSASILVAAAQWMDSVASDLFSAASAFQSVVWG 60
Qy 61 LTVGSIWGSAGGLMVAASAPYVAMSVTAQAEITAAQVRVAAAAYETAYGLTVPVPIA 120
Db 61 LTVGSIWGSAGGLMVAASAPYVAMSVTAQAEITAAQVRVAAAAYETAYGLTVPVPIA 120
Qy 121 ENRAELMILITATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180
Db 121 ENRAELMILITATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180
Qy 181 PEMTSAGGLLEQAAAEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYS 240
Db 181 PEMTSAGGLLEQAAAEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYS 240
Qy 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAQVCTAAQNGVRAMSSL 300
Db 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAQVCTAAQNGVRAMSSL 300
Qy 301 GSSILGSSGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAARERG 360
Db 301 GSSILGSSGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAARERG 360
Qy 361 QMLGGLPVGQMGARAGGGLSGVLRRPVPYMPHSPAAGDIAP----- 403
Db 361 QMLGGLPVGQMGARAGGGLSGVLRRPVPYMPHSPAAGDIAP----- 420
Qy 404 -PALSQDRFADFALPLDPS-----AMVAQVGPQVNVINTKLYNNNAVGAG 448
Db 421 LNATDPGAAQAFNAPVQAQSVLRNFLAAPPRAAQAQVCTAAQNGVRAMSSL 479
Qy 449 TGIVIDPENGVLTNHVIAGATDINAFSVGSGQTYGVDDVGVYDRTQDVAVLQIRGAGGLP 508
Db 480 CN-----NYELMTINYQFG---DVDA-----HGAMIRAQAASLEAHOAIVRDVLAAGDFW 527
Qy 509 SAAIGGVAVGEPVAMG-----NSGGQGTTPRAVGRVVALQOTVQASDLSLTGA 558
Db 528 GGA--GSVACQEFITQLGRNFPQVIYEQANAHQ-----KYQAGNNNAQOTDSAVGS 576
Qy 559 E-ETLNG 564
Db 577 SWATSN 583

RESULT 5
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-107

Query Match 64.8%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
DB 1 MVDGALPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLMVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
DB 61 SSAGLMVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
QY 129 LIATNLGQNTPAIVNEAEYGVEMWQDAAMFGYAAATATATATLLPPEEAPMTSAGG 188
DB 121 LIATNLGQNTPAIVNEAEYGVEMWQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
QY 189 LLEQAAVEEASDTPAAANQLMNNVPQALQQAQTOGTTTSSKGLGWLKTVSPHRSPISN 248
DB 181 LLEQAAVEEASDTPAAANQLMNNVPQALQQAQTOGTTTSSKGLGWLKTVSPHRSPISN 240
QY 249 MVSMAHHMNTSGVSMNTTSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 308
DB 241 MVSMAHHMNTSGVSMNTTSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 300
QY 309 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 369 GQMGARAGGGLSGVLRVPPRPYMPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRPYMPHSPAAG 391

RESULT 6
US-08-818-111-102
Sequence 102, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Hedrick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-102

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
DB 1 MVDGALPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLMVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
DB 61 SSAGLMVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
QY 129 LIATNLGQNTPAIVNEAEYGVEMWQDAAMFGYAAATATATATLLPPEEAPMTSAGG 188
DB 121 LIATNLGQNTPAIVNEAEYGVEMWQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
QY 189 LLEQAAVEEASDTPAAANQLMNNVPQALQQAQTOGTTTSSKGLGWLKTVSPHRSPISN 248
DB 181 LLEQAAVEEASDTPAAANQLMNNVPQALQQAQTOGTTTSSKGLGWLKTVSPHRSPISN 240
QY 249 MVSMAHHMNTSGVSMNTTSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 308
DB 241 MVSMAHHMNTSGVSMNTTSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 300
QY 309 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 369 GQMGARAGGGLSGVLRVPPRPYMPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRPYMPHSPAAG 391

RESULT 7
US-09-056-556-107
Sequence 107, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

TUBERCULOSIS

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-107

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
Db 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLWVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
Db 61 SSAGLWVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPFEAPEMTSAGG 188
Db 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
QY 189 LLEQAAVEEASDTAAANLNMNVPAQLQQLAQTQGTTPSSKLGGLWKTVPSPHSPISN 248
Db 181 LLEQAAVEEASDTAAANLNMNVPAQLQQLAQTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 249 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 308
Db 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 300
QY 309 LGGVVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
Db 301 LGGVVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

RESULT 8
US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Nero, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-102

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
Db 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLWVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
Db 61 SSAGLWVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 188
Db 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
QY 189 LLEQAAVEEASDTAAANLNMNVPAQLQQLAQTQGTTPSSKLGGLWKTVPSPHSPISN 248
Db 181 LLEQAAVEEASDTAAANLNMNVPAQLQQLAQTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 249 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 308
Db 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 300
QY 309 LGGVVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
Db 301 LGGVVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

RESULT 9
US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Hendrickson, Ronald C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 355
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/072,967
 FILING DATE: 05-MAY-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 107:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 391 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-072-967-107

Query Match 64.8%; Score 1949; DB 4; Length 391;
 Best Local Similarity 100.0%; Pred. No. 9.4e-140; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0;

Qy	9	MVDFGALPPEINSARMYAGSGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG	68
Db	1	MVDFGALPPEINSARMYAGSGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG	60
Qy	69	SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI	128
Db	61	SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI	120
Qy	129	LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG	188
Db	121	LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG	180
Qy	189	LLEQAAVEAEADTAANQLMNVFQALQQAQPTGGTTPSSKLGGLWKTVPSPHRSPISN	248
Db	181	LLEQAAVEAEADTAANQLMNVFQALQQAQPTGGTTPSSKLGGLWKTVPSPHRSPISN	240
Qy	249	MYSMANNHSMNTNSGVSMNTNLSMLKGFAPAAAACAVQTAQNGVRAMSSIGSLGSSG	308
Db	241	MYSMANNHSMNTNSGVSMNTNLSMLKGFAPAAAACAVQTAQNGVRAMSSIGSLGSSG	300
Qy	309	LCGGVAANLGRAASVGSISVPOQAAAAQVTPAARALPLTSLTSAABERGPGQMLGGLPV	368
Db	301	LCGGVAANLGRAASVGSISVPOQAAAAQVTPAARALPLTSLTSAABERGPGQMLGGLPV	360
Qy	369	GOMGARAGGLSVLRVPRPYVMPHSFPAAG	399
Db	361	GOMGARAGGLSVLRVPRPYVMPHSFPAAG	391

RESULT 10
 US-08-818-112-111
 ; Sequence 111, Application US/08818112
 ; Patent No. 6290369
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/818,112
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 111:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 396 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-08-818-112-111

Query Match 55.0%; Score 1652.5; DB 3; Length 396;
 Best Local Similarity 84.9%; Pred. No. 2.7e-117; Indels 7; Gaps 3;
 Matches 337; Conservative 19; Mismatches 34;

Qy	9	MVDFGALPPEINSARMYAGSGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG	68
Db	1	MVDFGALPPEINSARMYAGSGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG	60
Qy	69	SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI	128
Db	61	SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI	120
Qy	129	LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG	188
Db	121	LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEAPEMTSAGG	180
Qy	189	LLEQAAVEAEADTAANQLMNVFQALQQAQPTGGTTPSSKLGGLWKTVPSPHRSPISN	248
Db	181	LLEQAAVEAEADTAANQLMNVFQALQQAQPTGGTTPSSKLGGLWKTVPSPHRSPISN	240
Qy	249	MYSMANNHSMNTNSGVSMNTNLSMLKGFAPAAAACAVQTAQNGVRAMSSIGSLGSSG	304
Db	241	MYSMANNHSMNTNSGVSMNTNLSMLKGFAPAAAACAVQTAQNGVRAMSSIGSLGSSG	299
Qy	305	GSSGLGGGVAANLGRAASVGSISVPOQAAAAQVTPAARALPLTSLTSAABERGPGQMLG	364

Db 300 GSSGLGAGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAQAOTAPGHMLG 359
QY 365 GLPVQGMGARAG--GGLSGVLRVPPRPYPVMPHSPAAG 399
Db 360 GLPLGQLTNSGGFGVSNALRMPRAVYVMPRVPAAG 396
RESULT 11
US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Reto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-111-106
Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 9 MVDFGALPPEINSAARMYAGPGSASLVAAQAQWDSVADLFSAAAFQSVVWGLTVGWSWG 68
Db 1 VVDFGALPPEINSAARMYAGPGSASLVAAQAQWDSVADLFSAAAFQSVVWGLTVGWSWG 60
QY 69 SSAGLMVAASPSYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128
Db 61 SSAGLMVAASPSYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 188
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEDAPLIINPGG 180
QY 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTGQTPSSKLGGLWKTVPSPHRSPISN 248
Db 181 LLEQAAVEEAIDTAAANQLMNNVPQALQQAQPTKSIWPFQDLSLWKAISPHLSPLSN 240
QY 249 MVSMANNHMTNSGVSMWTNTLSSMLKGFAPAAAAQAQVTAQAQNGVRAKSS-----LGSSSL 304

Db 241 IVSMLNNHVSMTNSGVSMASTLHSLMLKGFAP-AAQAQVETAAQNGVQAMSSLSGSLGSSSL 299
QY 305 GSSGLGAGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGPGOMLG 364
Db 300 GSSGLGAGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAQAOTAPGHMLG 359
QY 365 GLPVQGMGARAG--GGLSGVLRVPPRPYPVMPHSPAAG 399
Db 360 GLPLGQLTNSGGFGVSNALRMPRAVYVMPRVPAAG 396
RESULT 12
US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-111
Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 9 MVDFGALPPEINSAARMYAGPGSASLVAAQAQWDSVADLFSAAAFQSVVWGLTVGWSWG 68
Db 1 VVDFGALPPEINSAARMYAGPGSASLVAAQAQWDSVADLFSAAAFQSVVWGLTVGWSWG 60
QY 69 SSAGLMVAASPSYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128
Db 61 SSAGLMVAASPSYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 188
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEDAPLIINPGG 180
QY 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTGQTPSSKLGGLWKTVPSPHRSPISN 248
Db 181 LLEQAAVEEAIDTAAANQLMNNVPQALQQAQPTKSIWPFQDLSLWKAISPHLSPLSN 240

QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSS----LGSSL 304
DB 241 IVSMLNNHVSMTNSGVSMNTLSSMLKGFAP-AAAQAVETAQNGVQAMSSLSGSLGSSL 299
QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTFAAALPLTSLTSAERGPQMLG 364
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANQAVTFAAALPLTSLTSAAGPAGHMLG 359
QY 365 GLPVQMGARAG--GGLSGVLRVPRPVYVMPHSPAAG 399
DB 360 GLPLGQLTNSGGFGGVSNALRMPPRAYVMPRVPAAAG 396

RESULT 13
US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-106

Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 1 VVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIG 60

QY 69 SSAGLWAAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 128
DB 61 SSAGLWAAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120

QY 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAATAATATALLPFEAPWTSAGG 188

DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAATAATATALLPFEAPLITNPGG 180
QY 189 LLEQAAVEEAASDTAAANQLMNNVPQALQQLAQPTGGTTPSSKLGGLWKTVPSPHRSPISN 248
DB 181 LLEQAVAVEEAIDTAAANQLMNNVPQALQQLAQPTKSIWPFQJSELWKAISPHLSPLSN 240
QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSS----LGSSL 304
DB 241 IVSMLNNHVSMTNSGVSMNTLSSMLKGFAP-AAAQAVETAQNGVQAMSSLSGSLGSSL 299
QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTFAAALPLTSLTSAERGPQMLG 364
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANQAVTFAAALPLTSLTSAAGPAGHMLG 359
QY 365 GLPVQMGARAG--GGLSGVLRVPRPVYVMPHSPAAG 399
DB 360 GLPLGQLTNSGGFGGVSNALRMPPRAYVMPRVPAAAG 396

RESULT 14
US-09-072-967-111
; Sequence 111, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-967-111

Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIG 68

Db 1 VVDFGALPPEINSARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTTGWSWIG 60
Qy 69 SSAGLMVAASAPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128
Db 61 SSAGLMVAASAPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
Qy 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLIPFEAPEMTSAGG 188
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLIPFEAPITNPGG 180
Qy 189 LLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTSPHRSPTSN 248
Db 181 LLEQAAVVEEADITAAANQLMNNVPOALQQLAQPTKSIWPDQLSELWKAISPHLSPLSN 240
Qy 249 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVCTAAQNGVRAMSS-----LGSSL 304
Db 241 IVSMNLNHVMTNSGVSMNTLSSMLKGFAP-AAAQAVETAQNGVQAVSSLGSLGSSL 299
Qy 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAAEERGPGOMLG 364
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAAQTAPEGHMLG 359
Qy 365 GLFVQMGARAG--GGLSGVLVPRPPYMPHSPAG 399
Db 360 GLFLGQLTNSGGFGGVSNAIRMPPRAYMPRPVPAAG 396

RESULT 15

US-08-818-112-109
; Sequence 109, Application US/09818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF INVENTION: 153
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-818-112-109

Query Match 49.4%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 8.8e-105;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;
Qy 9 MVDFGALPPEINSARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTTGWSWIG 68
Db 1 VVDFGALPPEINSARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTTGWSWIG 60
Qy 69 SSAGLMVAASAPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128
Db 61 SSAGLMVAASAPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMT 120
Qy 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLIPFEAPEMTSAGG 188
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLIPFEAPITNPGG 180
Qy 189 LLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTSPHRSPTSN 248
Db 181 LLEQAAVVEEADITAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTAVSPHLSPLSN 240
Qy 249 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVCTAAQNGVRAMSS-----LGSSL 304
Db 241 VSSIANNHMTNSGVSMNTLSSMLKGFAP-AAAQAVETAENGWVWAMSSLGSLGSSL 299
Qy 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAAEERGPGOMLG 364
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPPAWAANAQVTPAARALPLTSLTSAAQTAPEGHMLG 359

Search completed: June 22, 2004, 17:27:36

Job time : 22.2754 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 55.2012 seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-20

Perfect score: 3007

Sequence: 1 HMMHHHHMDFGALPPBEIN.....SGGPVNLGQVGMNTAAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3007	100.0	596	9	US-09-287-849-26
2	3007	100.0	596	12	US-09-886-349A-20
3	3007	100.0	596	14	US-10-359-460-26
4	3007	100.0	596	14	US-10-098-732A-20
5	2946	98.0	729	15	US-10-369-983-21
6	2946	98.0	813	15	US-10-369-983-15
7	2946	98.0	825	15	US-10-369-983-14
8	2946	98.0	875	15	US-10-369-983-13
9	2946	98.0	930	14	US-10-098-732A-65
10	2946	98.0	930	15	US-10-369-983-12
11	2946	98.0	1016	15	US-10-369-983-18
12	2946	98.0	1022	15	US-10-369-983-17
13	2946	98.0	1154	15	US-10-369-983-16
14	2943	97.9	729	15	US-09-886-349A-18
15	2943	97.9	729	14	US-10-098-732A-18

```
16 2943 97.9 729 15 US-10-369-983-22 Sequence 22, Appl
17 2941 97.8 729 9 US-09-287-849-2 Sequence 2, Appli
18 2941 97.8 729 12 US-09-886-349A-16 Sequence 16, Appl
19 2941 97.8 729 14 US-10-359-460-2 Sequence 2, Appli
20 2941 97.8 729 14 US-10-098-732A-16 Sequence 16, Appl
21 2941 97.8 729 15 US-10-359-459-2 Sequence 2, Appli
22 2033.5 67.6 600 9 US-09-287-849-22 Sequence 22, Appl
23 2033.5 67.6 600 14 US-10-359-460-22 Sequence 22, Appl
24 1955 65.0 1010 15 US-10-369-983-4 Sequence 4, Appli
25 1949 64.8 391 12 US-09-886-349A-14 Sequence 14, Appli
26 1949 64.8 391 14 US-10-193-002-102 Sequence 102, App
27 1949 64.8 391 14 US-10-084-843-107 Sequence 107, App
28 1949 64.8 391 14 US-10-098-732A-14 Sequence 14, Appli
29 1949 64.8 723 15 US-10-369-983-2 Sequence 8, Appli
30 1934 64.3 391 12 US-09-872-186-8 Sequence 106, App
31 1652.5 55.0 396 14 US-10-193-002-106 Sequence 111, App
32 1652.5 55.0 396 14 US-10-084-843-111 Sequence 62455, A
33 1583 52.6 393 12 US-10-282-122A-62455 Sequence 62455, A
34 1583 52.6 393 12 US-10-282-122A-64892 Sequence 64892, A
35 1486.5 49.4 359 14 US-10-193-002-104 Sequence 104, App
36 1486.5 49.4 359 14 US-10-084-843-109 Sequence 109, App
37 1189 39.5 358 9 US-09-287-849-8 Sequence 8, Appli
38 1189 39.5 358 14 US-10-359-460-8 Sequence 8, Appli
39 1187 39.5 263 12 US-09-886-349A-12 Sequence 12, Appli
40 1187 39.5 263 14 US-10-193-002-92 Sequence 92, Appli
41 1187 39.5 263 14 US-10-084-843-91 Sequence 91, Appli
42 1187 39.5 263 14 US-10-098-732A-12 Sequence 12, Appli
43 989 32.9 355 9 US-09-712-363-161 Sequence 161, App
44 987 32.8 330 12 US-09-886-349A-4 Sequence 4, Appli
45 987 32.8 330 14 US-10-098-732A-4 Sequence 4, Appli
```

ALIGNMENTS

```
RESULT 1
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26
```

Query Match 100.0%; Score 3007; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;

Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HHHHHHHHVDGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWG 60
DB 1 HHHHHHHHVDGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSMTGSSAGLMVAASPVVAMMSVTAGAEELTAQVRAAAAYETAYGLTVPPVIA 120
DB 61 LTVGSMTGSSAGLMVAASPVVAMMSVTAGAEELTAQVRAAAAYETAYGLTVPPVIA 120

QY 121 ENRAELMILIAATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180
DB 121 ENRAELMILIAATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180

QY 181 PEMSAGGLLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGTTTSSKLGGLWKTYS 240
DB 181 PEMSAGGLLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGTTTSSKLGGLWKTYS 240

QY 241 PHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
DB 241 PHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOQMAAANQAVTPAARALPLTSLTSAABERGPG 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOQMAAANQAVTPAARALPLTSLTSAABERGPG 360

QY 361 QMLGGLPVGQMGARAGGSLGVLRVPPRYVMPHSPAAAGDIAPPALSDQREDFPALPLD 420
DB 361 QMLGGLPVGQMGARAGGSLGVLRVPPRYVMPHSPAAAGDIAPPALSDQREDFPALPLD 420

QY 421 PSAMVAQVGPQVYNINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVYNINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480

QY 481 QTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
DB 481 QTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTCABETLNLQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 541 RVVALGQTVQASDSLTCABETLNLQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596

RESULT 2

US-09-886-349A-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/866,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MT559F)
US-09-886-349A-20

Query Match

Best Local Similarity 100.0%; Score 3007; DB 12; Length 596;

; Pred. No. 5e-202;

Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HHHHHHHHVDGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWG 60
DB 1 HHHHHHHHVDGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSMTGSSAGLMVAASPVVAMMSVTAGAEELTAQVRAAAAYETAYGLTVPPVIA 120
DB 61 LTVGSMTGSSAGLMVAASPVVAMMSVTAGAEELTAQVRAAAAYETAYGLTVPPVIA 120

QY 121 ENRAELMILIAATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180
DB 121 ENRAELMILIAATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180

QY 181 PEMSAGGLLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGTTTSSKLGGLWKTYS 240
DB 181 PEMSAGGLLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGTTTSSKLGGLWKTYS 240

QY 241 PHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
DB 241 PHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOQMAAANQAVTPAARALPLTSLTSAABERGPG 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOQMAAANQAVTPAARALPLTSLTSAABERGPG 360

QY 361 QMLGGLPVGQMGARAGGSLGVLRVPPRYVMPHSPAAAGDIAPPALSDQREDFPALPLD 420
DB 361 QMLGGLPVGQMGARAGGSLGVLRVPPRYVMPHSPAAAGDIAPPALSDQREDFPALPLD 420

QY 421 PSAMVAQVGPQVYNINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVYNINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480

QY 481 QTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
DB 481 QTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTCABETLNLQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 541 RVVALGQTVQASDSLTCABETLNLQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596

RESULT 3

US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1


```
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TdH9-Ra35 (designated MTB59F)
US-10-359-460-26

Query Match      100.0%; Score 3007; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHHMDVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHHHMDVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWISSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120
DB 61 LTVGSWISSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120

QY 61 LTVGSWISSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120
DB 61 LTVGSWISSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120

QY 121 ENRAELMILIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPPEEA 180
DB 121 ENRAELMILIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPPEEA 180

QY 181 PENTSAGGLEQAAAEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVS 240
DB 181 PENTSAGGLEQAAAEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVS 240

QY 241 PHRSPIINMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
DB 241 PHRSPIINMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300

QY 301 GSSLSGSLGGVGAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGP 360
DB 301 GSSLSGSLGGVGAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGP 360

QY 361 QMLGGLPVQMGARAGGSLGVLRPVPPVPHSPAGDIAPPALSODRFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGSLGVLRPVPPVPHSPAGDIAPPALSODRFADFPALPLD 420

QY 421 PSAMVAQVGPQVNVINTKLYNNVAVGAGTGVDPNGVVLTNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVNVINTKLYNNVAVGAGTGVDPNGVVLTNHVIAGATDINAFSVGSG 480

QY 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540
DB 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
```

RESULT 4

```
US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; NUMBER OF SEQ ID NOS: 837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TdH9-Ra35 (designated MTB59F)
US-10-098-732A-20

Query Match      100.0%; Score 3007; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHHMDVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHHHMDVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWISSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120
DB 61 LTVGSWISSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120

QY 61 LTVGSWISSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120
DB 61 LTVGSWISSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120

QY 121 ENRAELMILIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPPEEA 180
DB 121 ENRAELMILIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPPEEA 180

QY 181 PENTSAGGLEQAAAEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVS 240
DB 181 PENTSAGGLEQAAAEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVS 240

QY 241 PHRSPIINMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
DB 241 PHRSPIINMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300

QY 301 GSSLSGSLGGVGAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGP 360
DB 301 GSSLSGSLGGVGAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGP 360

QY 361 QMLGGLPVQMGARAGGSLGVLRPVPPVPHSPAGDIAPPALSODRFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGSLGVLRPVPPVPHSPAGDIAPPALSODRFADFPALPLD 420

QY 421 PSAMVAQVGPQVNVINTKLYNNVAVGAGTGVDPNGVVLTNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVNVINTKLYNNVAVGAGTGVDPNGVVLTNHVIAGATDINAFSVGSG 480

QY 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540
DB 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
```

RESULT 5

```
US-10-369-983-21
; Sequence 21, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F
US-10-369-983-21

```

Query Match	98.0%; Score 2946; DB 15; Length 729;
Best Local Similarity	100.0%; Pred. No. 1.2e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	9 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVADLFSAAAFOSVVMGLTVGSWIG 68
Db	142 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVADLFSAAAFOSVVMGLTVGSWIG 201
Qy	69 SSAGLMVAASPYVWNSVTVAGQBELTAAQVRVAAAAAYETAYGLTVPPVIAENRAELMI 128
Db	202 SSAGLMVAASPYVWNSVTVAGQBELTAAQVRVAAAAAYETAYGLTVPPVIAENRAELMI 261
Qy	129 LIATNLGLQNTPAIVANAEYBGEWMAQDAAMFCYAAATATATATLPPFEAPEMTSAGG 188
Db	262 LIATNLGLQNTPAIVANAEYBGEWMAQDAAMFCYAAATATATATLPPFEAPEMTSAGG 321
Qy	189 LLQQAARAVEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVSPHRSPI 248
Db	322 LLQQAARAVEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVSPHRSPI 381
Qy	249 MVSMANNHMTNSGVSMNTLTLSMLKGFAPAAAAQVQTAQNGVRAMSLGSSLGSSG 308
Db	382 MVSMANNHMTNSGVSMNTLTLSMLKGFAPAAAAQVQTAQNGVRAMSLGSSLGSSG 441
Qy	309 LGGGVAANLGRASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGGQWMLGGLPV 368
Db	442 LGGGVAANLGRASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGGQWMLGGLPV 501
Qy	369 GQMGARAGGSLGVLRVPPRPYVWPHSPAGDITAPPALQDRPADFPALPLDPSAMVAQV 428
Db	502 GQMGARAGGSLGVLRVPPRPYVWPHSPAGDITAPPALQDRPADFPALPLDPSAMVAQV 561
Qy	429 GPQVWNINTKLGYNNAVCGAGTGIVDPNGVVLTNNHVIAGATDINAFSVSGSGQTYGVDV 488
Db	562 GPQVWNINTKLGYNNAVCGAGTGIVDPNGVVLTNNHVIAGATDINAFSVSGSGQTYGVDV 621
Qy	489 GYDRTQDVAVLQIRGAGGLPSAII GGGVAVGEPVVMVAMNSGGQGTTPRVPGRVVALQGT 548
Db	622 GYDRTQDVAVLQIRGAGGLPSAII GGGVAVGEPVVMVAMNSGGQGTTPRVPGRVVALQGT 681
Qy	549 VQASDSLTAAGTEELNGLIQFDAAIQPGDSGGPVYVNGLQTVGMNTAAS 596
Db	682 VQASDSLTAAGTEELNGLIQFDAAIQPGDSGGPVYVNGLQTVGMNTAAS 729

```

RESULT 6
US-10-369-983-15
; Sequence 15, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-0090811US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DPV)

```

Query Match 98.0%; Score 2946; DB 15; Length 825;
Best Local Similarity 100.0%; Pred. No. 1.4e-137; Indels 0; Gaps 0;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPPEIN SARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 142 MYDFGALPPEIN SARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLWVAASPYVAMSVTACQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
DB 202 SSAGLWVAASPYVAMSVTACQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 261

QY 129 LIATNLLGONTPTAI AVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 188
DB 262 LIATNLLGONTPTAI AVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 321

QY 189 LLEQAAAVEREASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPTSN 248
DB 322 LLEQAAAVEREASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPTSN 381

QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 308
DB 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 441

QY 309 LGGGVAANLGRAASVGSLSVPOQAAAAQAVTPAARALPLTSLTSAABERGQMLGGLPV 368
DB 442 LGGGVAANLGRAASVGSLSVPOQAAAAQAVTPAARALPLTSLTSAABERGQMLGGLPV 501

QY 369 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAAGDIAPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAAGDIAPALSDQRFADFPALPLDPSAMVAQV 561

QY 429 GPQVNVINTKLGYNNAVAGTGVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVDV 488
DB 562 GPQVNVINTKLGYNNAVAGTGVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVDV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVAVMGNSGGGQTPRAVPGRVVVALGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVAVMGNSGGGQTPRAVPGRVVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 729

RESULT 8
US-10-369-983-13
; Sequence 13, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MT89F (MTB72F-Brd14)
US-10-369-983-13

Query Match 98.0%; Score 2946; DB 15; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.5e-197;

Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPPEIN SARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 142 MYDFGALPPEIN SARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLWVAASPYVAMSVTACQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
DB 202 SSAGLWVAASPYVAMSVTACQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 261

QY 129 LIATNLLGONTPTAI AVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 188
DB 262 LIATNLLGONTPTAI AVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 321

QY 189 LLEQAAAVEREASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPTSN 248
DB 322 LLEQAAAVEREASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPTSN 381

QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 308
DB 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 441

QY 309 LGGGVAANLGRAASVGSLSVPOQAAAAQAVTPAARALPLTSLTSAABERGQMLGGLPV 368
DB 442 LGGGVAANLGRAASVGSLSVPOQAAAAQAVTPAARALPLTSLTSAABERGQMLGGLPV 501

QY 369 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAAGDIAPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAAGDIAPALSDQRFADFPALPLDPSAMVAQV 561

QY 429 GPQVNVINTKLGYNNAVAGTGVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVDV 488
DB 562 GPQVNVINTKLGYNNAVAGTGVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVDV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVAVMGNSGGGQTPRAVPGRVVVALGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVAVMGNSGGGQTPRAVPGRVVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 729

RESULT 9
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: US 60/275,837
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Ra12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match 98.0%; Score 2946; DB 14; Length 930;

Best Local Similarity 100.0%; Pred. No. 1.6e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLMTVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
DB 202 SSAGLMTVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261

QY 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPPEAPEMTSAGG 188
DB 262 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGG 321

QY 189 LLEQAAAVEASDTAAANQLMNVPOALQOLAQTOGTTPSSKLGGLWKTSPHRSPIIN 248
DB 322 LLEQAAAVEASDTAAANQLMNVPOALQOLAQTOGTTPSSKLGGLWKTSPHRSPIIN 381

QY 249 MYSMANNHMTNSGYSMVTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSLGSSG 308
DB 382 MYSMANNHMTNSGYSMVTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSLGSSG 441

QY 309 LGGGVAANLGRAASVGSLSVPOQAWAANAQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 442 LGGGVAANLGRAASVGSLSVPOQAWAANAQAVTPAARALPLTSLTSAERGPQMLGGLPV 501

QY 369 GQMGARAGGGLSVLRVPPRYVPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSVLRVPPRYVPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV 561

QY 429 GPQVNNITKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVV 488
DB 562 GPQVNNITKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAMGNSGGQGTTPRAVPGRVVAGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAMGNSGGQGTTPRAVPGRVVAGQT 681

QY 549 VQASDSLTGAETLNGLIQFDAAIQPDGSGPPVNGLGQVGMNTAAS 596
DB 682 VQASDSLTGAETLNGLIQFDAAIQPDGSGPPVNGLGQVGMNTAAS 729

RESULT 10

US-10-369-983-12
; Sequence 12, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: R95F (MTB72F-MAPS)
US-10-369-983-12

Query Match 98.0%; Score 2946; DB 15; Length 930;
Best Local Similarity 100.0%; Pred. No. 1.6e-197;

QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLMTVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
DB 202 SSAGLMTVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261

QY 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGG 188
DB 262 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGG 321

QY 189 LLEQAAAVEASDTAAANQLMNVPOALQOLAQTOGTTPSSKLGGLWKTSPHRSPIIN 248
DB 322 LLEQAAAVEASDTAAANQLMNVPOALQOLAQTOGTTPSSKLGGLWKTSPHRSPIIN 381

QY 249 MYSMANNHMTNSGYSMVTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSLGSSG 308
DB 382 MYSMANNHMTNSGYSMVTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSLGSSG 441

QY 309 LGGGVAANLGRAASVGSLSVPOQAWAANAQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 442 LGGGVAANLGRAASVGSLSVPOQAWAANAQAVTPAARALPLTSLTSAERGPQMLGGLPV 501

QY 369 GQMGARAGGGLSVLRVPPRYVPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSVLRVPPRYVPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV 561

QY 429 GPQVNNITKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVV 488
DB 562 GPQVNNITKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAMGNSGGQGTTPRAVPGRVVAGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAMGNSGGQGTTPRAVPGRVVAGQT 681

QY 549 VQASDSLTGAETLNGLIQFDAAIQPDGSGPPVNGLGQVGMNTAAS 596
DB 682 VQASDSLTGAETLNGLIQFDAAIQPDGSGPPVNGLGQVGMNTAAS 729

RESULT 11

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85B)
US-10-369-983-18

Query Match 98.0%; Score 2946; DB 15; Length 1016;
Best Local Similarity 100.0%; Pred. No. 1.8e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	142	MYDFGLPPEINSAEYAGPSASIVAAQWSDVSASDLFSAASAFQSVVWGLTVGSWIG	201
Qy	69	SSAGLWVAASPYVAMGVTAQAEELTAQVVRVAAAAAYETAYGLTVPPPVIAENRAELMI	128
Db	202	SSAGLWVAASPYVAMGVTAQAEELTAQVVRVAAAAAYETAYGLTVPPPVIAENRAELMI	261
Qy	129	LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATALLPEEAPEMTSAGG	188
Db	262	LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATALLPEEAPEMTSAGG	321
Qy	189	LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISN	248
Db	322	LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISN	381
Qy	249	MYSMANNHSMWNTSGVSTNTLSSMLKGFAPAAQAQAVQTAQAQNGVRAMSSILGSSILGSSG	308
Db	382	MYSMANNHSMWNTSGVSTNTLSSMLKGFAPAAQAQAVQTAQAQNGVRAMSSILGSSILGSSG	441
Qy	309	LGCGVAAANLGRAASVGSLSVPOWAAAAANQAVTPPAARALPLTSLTSAAEKRGFGMLGGLPV	368
Db	442	LGCGVAAANLGRAASVGSLSVPOWAAAAANQAVTPPAARALPLTSLTSAAEKRGFGMLGGLPV	501
Qy	369	GQMGARAGGILSGVLRVPRPYWPHSPAAQDIAAPPALSQDRFADFPALPLDPSAMVAQV	428
Db	502	GQMGARAGGILSGVLRVPRPYWPHSPAAQDIAAPPALSQDRFADFPALPLDPSAMVAQV	561
Qy	429	GQVQVNIINTKLGYNNAVAGAGTGIVIDPNGVVLTNHVIAGATDINAFSVSGGQTYGVVV	488
Db	562	GQVQVNIINTKLGYNNAVAGAGTGIVIDPNGVVLTNHVIAGATDINAFSVSGGQTYGVVV	621
Qy	489	GYDRTQDVAVLQLRAGGLPSSAATGGGVAVGEPVWAMGNSGGGGTTPRAVPGRVVAGQT	548
Db	622	GYDRTQDVAVLQLRAGGLPSSAATGGGVAVGEPVWAMGNSGGGGTTPRAVPGRVVAGQT	681
Qy	549	VQASDSLTAAGETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMNTAAS	596
Db	682	VQASDSLTAAGETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMNTAAS	729

RESULT 12

US-10-369-983-17

; Sequence 17, Application US/10369983

; Publication No. US20030235593A1

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis

; FILE REFERENCE: 014058-009081US

; CURRENT APPLICATION NUMBER: US/10/369,983

; CURRENT FILING DATE: 2003-02-18

; PRIOR APPLICATION NUMBER: US 60/357,351

; PRIOR FILING DATE: 2002-02-15

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 17

; LENGTH: 1022

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: fusion protein

; OTHER INFORMATION: MTB102tm2F (MTB102FTM2, MTB72F-hTCC#1)

; US-10-369-983-17

Query Match	98.0%;	Score 2946;	DB 15;	Length 1022;
Best Local Similarity	100.0%;	Pred. No. 1.8e-197;		
Matches 588;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
Qy	9	MVDFGALPPEINSARMYAGPSASILVAAQWDSVAGDLFSASAFQSVVMGTVGSGWG	68	
Db	142	MVDFGALPPEINSARMYAGPSASILVAAQWDSVAGDLFSASAFQSVVMGTVGSGWG	201	

QY	69	SSAGLWVAARSPYVAMSVTAGQAELETAQVTVAAAAAYETAYGILTVPPPVIAENRAELMI	128
Db	202	SSAGLWVAARSPYVAMSVTAGQAELETAQVTVAAAAAYETAYGILTVPPPVIAENRAELMI	261
QY	129	LIATNLLGQNTPAI VNAEAYGEMWADAAAFYGAATAATATATATLLPFEAPEMTSGG	188
Db	262	LIATNLLGQNTPAI VNAEAYGEMWADAAAFYGAATAATATATATLLPFEAPEMTSGG	321
QY	189	LLSQAAVEEASDTAAANOLMNVPQALQOALOQTOGTTTSSKLGGLWKTVPHRSPLSN	248
Db	322	LLSQAAVEEASDTAAANOLMNVPQALQOALOQTOGTTTSSKLGGLWKTVPHRSPLSN	381
QY	249	MVSMANNHMTNSGVSMNTLTSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSLGSSG	308
Db	382	MVSMANNHMTNSGVSMNTLTSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSLGSSG	441
QY	309	LGGSVAANLGRASVGSLSVPOAWAAANQVTPAARALPLTSLTSAERGGQMLGGLPV	368
Db	442	LGGSVAANLGRASVGSLSVPOAWAAANQVTPAARALPLTSLTSAERGGQMLGGLPV	501
QY	369	GQMGARAGGSLGVLRVPPRPYVMPHSPAAGDIAPPALSODRFADFPALPLDPSAMVAQV	428
Db	502	GQMGARAGGSLGVLRVPPRPYVMPHSPAAGDIAPPALSODRFADFPALPLDPSAMVAQV	561
QY	429	GPQVNVINTKLGYNNAVGAAGTIVDPNGVYLTNNHVIAGATDINAFSVGSGQTVGVDDV	488
Db	562	GPQVNVINTKLGYNNAVGAAGTIVDPNGVYLTNNHVIAGATDINAFSVGSGQTVGVDDV	621
QY	489	GYDRTQDVAVLQLRGAGLPSAATGGGVAYGEPVWAMNSGGQGTTPRAVGRVVVALGQT	548
Db	622	GYDRTQDVAVLQLRGAGLPSAATGGGVAYGEPVWAMNSGGQGTTPRAVGRVVVALGQT	681
QY	549	VQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVYVGMNTAAS	596
Db	682	VQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVYVGMNTAAS	729
RESULT 13			
US-10-369-983-16			
; Sequence 16, Application US/10369983			
; Publication No. US20030235593A1			
; GENERAL INFORMATION:			
; APPLICANT: Skeiky, Yasir			
; APPLICANT: Guderian, Jeff			
; APPLICANT: Reed, Steven			
; APPLICANT: Corixa Corporation			
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis			
; FILE REFERENCE: 014058-009081US			
; CURRENT APPLICATION NUMBER: US/10/369,983			
; CURRENT FILING DATE: 2003-02-18			
; PRIOR APPLICATION NUMBER: US 60/357,351			
; PRIOR FILING DATE: 2002-02-15			
; NUMBER OF SEQ ID NOS: 22			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 16			
; LENGTH: 1154			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein			
; OTHER INFORMATION: MTB114F (MTB72p-mTCC#2)			
US-10-369-983-16			

		98.0%;	Score 2946;	DB 15;	Length 1154;
		Best Local Similarity 100.0%;	Prid.No. 2.2e-197;		
		Matches 588;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0
Qy	9	MVDFGALPPPEINSARMYAGPGSASLVAAAAQWDSVASDLFSAAGAFQSVMGLTVGSWGIG	68		
Db	142	MVDFGALPPPEINSARMYAGPGSASLVAAAAQWDSVASDLFSAAGAFQSVMGLTVGSWGIG	201		
Qv	69	SSAGLMTAAASPYYAMWSVTGAQDELTAQAQVRVAAAAYETAYGLTVPFPVTAENRAELMI	128		

202 SSAGLWVAASPYVAMWSVTAGQAELETAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 261
129 LIATNLLGQNTPAIAYNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAEPEMTSAGG 188
262 LIATNLLGQNTPAIAYNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAEPEMTSAGG 321
189 LLEQAAAVEEASDTAAANQNMNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI SN 248
322 LLEQAAAVEEASDTAAANQNMNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI SN 381
249 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAVMSLSSSLGSSG 308
382 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAVMSLSSSLGSSG 441
309 LGGGVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAAEERPGOMLGGPLV 368
442 LGGGVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAAEERPGOMLGGPLV 501
369 GQMGARAGGSLGVLVPRPYVMPHSPAAAGDIAPPALSDQDRFADFPALPLDPSAMVAQV 428
502 GQMGARAGGSLGVLVPRPYVMPHSPAAAGDIAPPALSDQDRFADFPALPLDPSAMVAQV 561
429 GPQVWNINIKLGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGSGTGYGVVV 488
562 GPQVWNINIKLGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGSGTGYGVVV 621
489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGNSGGGTPRAVPGRVVWALGOT 548
622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGNSGGGTPRAVPGRVVWALGOT 681
549 VQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 596
682 VQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 729

RESULT 14

US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Steven
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72FmutSA
; OTHER INFORMATION: (Ra12-TbHp-Ra35MutSA)
US-09-886-349A-18

Query Match 97.9%; Score 2943; DB 12; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.9e-197;
Matches 587; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
9 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 68
142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
69 SSAGLWVAASPYVAMWSVTAGQAELETAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128

202 SSAGLWVAASPYVAMWSVTAGQAELETAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 261
129 LIATNLLGQNTPAIAYNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAEPEMTSAGG 188
262 LIATNLLGQNTPAIAYNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAEPEMTSAGG 321
189 LLEQAAAVEEASDTAAANQNMNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI SN 248
322 LLEQAAAVEEASDTAAANQNMNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI SN 381
249 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAVMSLSSSLGSSG 308
382 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAVMSLSSSLGSSG 441
309 LGGGVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAAEERPGOMLGGPLV 368
442 LGGGVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAAEERPGOMLGGPLV 501
369 GQMGARAGGSLGVLVPRPYVMPHSPAAAGDIAPPALSDQDRFADFPALPLDPSAMVAQV 428
502 GQMGARAGGSLGVLVPRPYVMPHSPAAAGDIAPPALSDQDRFADFPALPLDPSAMVAQV 561
429 GPQVWNINIKLGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGSGTGYGVVV 488
562 GPQVWNINIKLGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGSGTGYGVVV 621
489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGNSGGGTPRAVPGRVVWALGOT 548
622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGNSGGGTPRAVPGRVVWALGOT 681
549 VQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 596
682 VQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 729

RESULT 15

US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72FmutSA
; OTHER INFORMATION: (Ra12-TbH9-Ra35MutSA)
US-10-098-732A-18

Query Match 97.9%; Score 2943; DB 14; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.9e-197;
Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
9 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 68
142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
69 SSAGLWVAASPYVAMWSVTAGQAELETAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128

Db	202	SSAGLWVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI	261
Qy	129	LIATNLLQONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG	188
Db	262	LIATNLLQONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG	321
Qy	189	LLEQAAAVEEASDTAAANCLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTVPHRSPISN	248
Db	322	LLEQAAAVEEASDTAAANCLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTVPHRSPISN	381
Qy	249	MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSSLSLSSGSSG	308
Db	382	MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSSLSLSSGSSG	441
Qy	309	LCGGVAANLGRAASVGSLSVPOAHHAAANOAVTPAARALPLTSLTSAABERGQMLGGLPV	368
Db	442	LCGGVAANLGRAASVGSLSVPOAHHAAANOAVTPAARALPLTSLTSAABERGQMLGGLPV	501
Qy	369	GOMGARAGGGLSGVLRVPPRPYVWPHSPAAAGDIAPPALSQDRFADFPALPLDPSAMVAQV	428
Db	502	GOMGARAGGGLSGVLRVPPRPYVWPHSPAAAGDIAPPALSQDRFADFPALPLDPSAMVAQV	561
Qy	429	GPQVNVINTKLYNNNAVAGAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVW	488
Db	562	GPQVNVINTKLYNNNAVAGAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVW	621
Qy	489	GYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGNSGGGTPRAVPGRVVALGQT	548
Db	622	GYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGNSGGGTPRAVPGRVVALGQT	681
Qy	549	VOASDSLTAABETLNGLIQFDAAIQFDSGGPVPVNGLQGVVGMNTAAS	596
Db	682	VOASDSLTAABETLNGLIQFDAAIQFDSGGPVPVNGLQGVVGMNTAAS	729

Search completed: June 22, 2004, 18:08:01
Job time : 56.2012 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 3.42692 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-22
Perfect score: 481
Sequence: 1 VAMSLTVAGVASADPVDVAV.....PGAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	481	100.0	96	3	US-08-818-112-102
2	481	100.0	96	4	US-08-818-111-97
3	481	100.0	96	4	US-09-056-556-102
4	481	100.0	96	4	US-09-072-596-97
5	481	100.0	96	4	US-09-072-967-102
6	481	100.0	110	4	US-09-050-739-151
7	419.5	87.2	331	4	US-09-223-040-4
8	419.5	87.2	331	4	US-09-287-843-4
9	419.5	87.2	368	4	US-09-287-849-24
10	419.5	87.2	433	4	US-09-287-849-14
11	419.5	87.2	856	4	US-09-287-849-12
12	419	87.1	299	4	US-09-287-849-19
13	419	87.1	600	4	US-09-287-849-22
14	419	87.1	710	4	US-09-287-849-16
15	153	32.2	30	4	US-10-005-920-5
16	132	27.4	26	4	US-10-005-920-4
17	125	26.0	24	4	US-10-005-920-1
18	121	25.2	24	4	US-10-005-920-2
19	107	22.2	22	4	US-10-005-920-3
20	104	21.6	20	3	US-08-818-112-120
21	104	21.6	20	4	US-08-818-111-115
22	104	21.6	20	4	US-09-056-556-120
23	104	21.6	20	4	US-09-072-596-115
24	104	21.6	20	4	US-09-072-967-120
25	93	19.3	20	3	US-08-818-112-54
26	93	19.3	20	4	US-08-818-111-54
27	93	19.3	20	4	US-09-056-556-54

28	93	19.3	20	4	US-09-072-596-54	Sequence 54, Appl
29	93	19.3	20	4	US-09-072-567-54	Sequence 54, Appl
30	74	15.4	1454	4	US-09-328-352-5793	Sequence 5793, Ap
31	73	15.2	626	2	US-08-596-300A-7	Sequence 7, Appl
32	73	15.2	626	2	US-08-596-300A-14	Sequence 14, Appl
33	72	15.0	314	1	US-08-040-753-2	Sequence 2, Appl
34	72	15.0	386	4	US-09-331-924-9	Sequence 9, Appl
35	72	15.0	386	4	US-10-147-951B-9	Sequence 9, Appl
36	72	15.0	393	4	US-09-331-924-8	Sequence 8, Appl
37	72	15.0	393	4	US-10-147-951B-8	Sequence 8, Appl
38	70.5	14.7	467	4	US-09-684-855-167	Sequence 167, App
39	70.5	14.7	541	4	US-09-252-991A-20524	Sequence 20524, A
40	70.5	14.7	1042	3	US-08-928-361B-11	Sequence 11, Appl
41	70.5	14.7	1042	4	US-09-588-995A-11	Sequence 11, Appl
42	70.5	14.7	1043	3	US-08-928-361B-30	Sequence 30, Appl
43	70.5	14.7	1721	3	US-08-700-651-5	Sequence 5, Appl
44	70.5	14.7	1721	3	US-08-928-361B-6	Sequence 6, Appl
45	70.5	14.7	1721	4	US-09-588-995A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-818-112-102
; Sequence 102, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-102

Query Match 100.0%; Score 481; DB 3; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.2e-50;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVAGSYLRNFLA 60
 Db 1 VAMSLTVGAGVADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVAGSYLRNFLA 60
 QY 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
 Db 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96

RESULT 2

US-08-818-111-97
 ; Sequence 97, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, David C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/818,111
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-6030
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 97:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 96 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-111-97

Query Match 100.0%; Score 481; DB 4; Length 96;
 Best Local Similarity 100.0%; Pred. No. 2.2e-50;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVAGSYLRNFLA 60
 Db 1 VAMSLTVGAGVADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVAGSYLRNFLA 60
 QY 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
 Db 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96

RESULT 3

US-09-056-556-102
 ; Sequence 102, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, David C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 ; NUMBER OF SEQUENCES: 241
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/056,556
 ; FILING DATE: 07-APR-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.457
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 102:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 96 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-056-556-102

Query Match 100.0%; Score 481; DB 4; Length 96;
 Best Local Similarity 100.0%; Pred. No. 2.2e-50;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVAGSYLRNFLA 60
 Db 1 VAMSLTVGAGVADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVAGSYLRNFLA 60
 QY 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
 Db 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96

RESULT 4

US-09-072-596-97
 ; Sequence 97, Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, David C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-97

Query Match 100.0%; Score 481; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.2e-50;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVQAQSYLRNFLA 60
DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVQAQSYLRNFLA 60

QY 61 APPQRAAMAAQLQAVPGAQYIGLVESVAGSCNNY 96
DB 61 APPQRAAMAAQLQAVPGAQYIGLVESVAGSCNNY 96

RESULT 5
US-09-072-967-102
Sequence 102, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESS: SEED AND BERRY LLP
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-102

Query Match 100.0%; Score 481; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.2e-50;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVQAQSYLRNFLA 60
DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVQAQSYLRNFLA 60

QY 61 APPQRAAMAAQLQAVPGAQYIGLVESVAGSCNNY 96
DB 61 APPQRAAMAAQLQAVPGAQYIGLVESVAGSCNNY 96

RESULT 6
US-09-050-739-151
Sequence 151, Application US/09050739
Patent No. 6641814
GENERAL INFORMATION:
APPLICANT: ANDERSEN, Peter
APPLICANT: NIELSEN, Rikke
APPLICANT: OETTINGER, Thomas
APPLICANT: RASMUSSEN, Peter Birk
APPLICANT: ROSENKRANDS, Ida
APPLICANT: WELDSINGH, Karin
APPLICANT: FLORIO, Walter
TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS
FILE REFERENCE: 670001-2002.1
CURRENT APPLICATION NUMBER: US/09/050,739
CURRENT FILING DATE: 1998-03-30
EARLIER APPLICATION NUMBER: 0376/97
EARLIER FILING DATE: 1997-04-02
EARLIER APPLICATION NUMBER: 1277/97
EARLIER FILING DATE: 1997-11-10
EARLIER APPLICATION NUMBER: 60/044,624
EARLIER FILING DATE: 1997-04-18
EARLIER APPLICATION NUMBER: 60/070,488
EARLIER FILING DATE: 1998-01-05
NUMBER OF SEQ ID NOS: 173
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 151
LENGTH: 110
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-050-739-151

Query Match 100.0%; Score 481; DB 4; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.6e-50;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVQAQSYLRNFLA 60
DB 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGQFNASPVQAQSYLRNFLA 74

QY 61 APPQRAAMAAQLQAVPGAQYIGLVESVAGSCNNY 96
DB 75 APPQRAAMAAQLQAVPGAQYIGLVESVAGSCNNY 110

RESULT 7
US-09-223-040-4
Sequence 4, Application US/09223040
Patent No. 6544522

```

; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-4

Query Match      87.2%; Score 419.5; DB 4; Length 331;
Best Local Similarity 80.4%; Pred. No. 2.8e-42;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 49
   |||
Db 129 LTVSVAVSEGKTEKHQIRSTNKLDPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 188
   |||

QY 50 VAQSYLRFNLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 96
   |||
Db 189 VAQSYLRFNLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 235
   |||

RESULT 8
US-09-287-849-4
; Sequence 4, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-4

Query Match      87.2%; Score 419.5; DB 4; Length 331;
Best Local Similarity 80.4%; Pred. No. 2.8e-42;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 49
   |||
Db 129 LTVSVAVSEGKTEKHQIRSTNKLDPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 188
   |||

QY 50 VAQSYLRFNLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 96
   |||
Db 189 VAQSYLRFNLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 235
   |||

RESULT 8
US-09-287-849-4
; Sequence 4, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-4

Query Match      87.2%; Score 419.5; DB 4; Length 331;
Best Local Similarity 80.4%; Pred. No. 2.8e-42;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

```

```

QY 5 LTVGAGVASA-----DPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 49
   |||
Db 129 LTVSVAVSEGKTEKHQIRSTNKLDPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 188
   |||

QY 50 VAQSYLRFNLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 96
   |||
Db 189 VAQSYLRFNLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 235
   |||

RESULT 9
US-09-287-849-24
; Sequence 24, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-24

Query Match      87.2%; Score 419.5; DB 4; Length 368;
Best Local Similarity 80.4%; Pred. No. 3.2e-42;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 49
   |||
Db 130 LTVSVAVSEGKTEKHQIRSTNKLDPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 189
   |||

QY 50 VAQSYLRFNLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 96
   |||
Db 190 VAQSYLRFNLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 236
   |||

RESULT 10
US-09-287-849-14
; Sequence 14, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849

```

```

QY      5  LTVGAGVASA-----DPDVAIVNTTCNGGVVAALNATDPGAAAQFNASP 49
      ||| : |||
Db     130 LTVSVAVSEGRPEKHIQIRSTNKLDPDVAIVNTTCNGGVVAALNATDPGAAAQFNASP 189
      ||| : |||
QY      50 VAQSYLRNFLAAPPPQRAAQAQVAPGAQAQYIGLVESVAGSCNNY 96
      ||| : |||
Db     190 VAQSYLRNFLAAPPPQRAAQAQVAPGAQAQYIGLVESVAGSCNNY 236
      ||| : |||

RESULT 12
US-09-287-849-19
; Sequence 19, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 19
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-19

Query Match      87.1%; Score 419; DB 4; Length 299;
Best Local Similarity 100.0%; Pred.No. 2.8e-42;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      15  DPDVAIVNTTCNGGVVAALNATDPGAAAQFNASPVQAQSYLRNFLAAPPPQRAAQAQIQ 74
      ||| : |||
Db     9  DPDVAIVNTTCNGGVVAALNATDPGAAAQFNASPVQAQSYLRNFLAAPPPQRAAQAQIQ 68
      ||| : |||
QY      75  AVPGAQAQYIGLVESVAGSCNNY 96
      ||| : |||
Db     69  AVPGAQAQYIGLVESVAGSCNNY 90
      ||| : |||

RESULT 13
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 19
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match      87.1%; Score 419; DB 4; Length 299;
Best Local Similarity 100.0%; Pred.No. 2.8e-42;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

```

; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-22

```

```

Query Match      87.1%; Score 419; DB 4; Length 600;
Best Local Similarity 100.0%; Pred. No. 7.1e-42;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 DPVDVAVINTTCNYGQVVAAALNATDPCGAAQFNASPVQAQSYLRNFLAAPPORAAQAQLQ 74
Db 402 DPVDVAVINTTCNYGQVVAAALNATDPCGAAQFNASPVQAQSYLRNFLAAPPORAAQAQLQ 461

```

```

Qy 75 AVPGAQYIGLVESVAGSCNNY 96
Db 462 AVPGAQYIGLVESVAGSCNNY 483

```

```

RESULT 14
US-09-287-849-16
; Sequence 16, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-16

```

```

Query Match      87.1%; Score 419; DB 4; Length 710;
Best Local Similarity 100.0%; Pred. No. 9e-42;

```

```

Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 DPVDVAVINTTCNYGQVVAAALNATDPCGAAQFNASPVQAQSYLRNFLAAPPORAAQAQLQ 74
Db 9 DPVDVAVINTTCNYGQVVAAALNATDPCGAAQFNASPVQAQSYLRNFLAAPPORAAQAQLQ 68

Qy 75 AVPGAQYIGLVESVAGSCNNY 96
Db 69 AVPGAQYIGLVESVAGSCNNY 90

RESULT 15
US-10-005-920-5
; Sequence 5, Application US/10005920
; Patent No. 6664096
; GENERAL INFORMATION:
; APPLICANT: Zhang, Ying
; TITLE OF INVENTION: Methods for Improved Diagnosis and Treatment of Mycobacterial Infection
; FILE REFERENCE: 03940034aa
; CURRENT APPLICATION NUMBER: US/10/005,920
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,785
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: US 60/294,602
; PRIOR FILING DATE: 2001-06-01
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-005-920-5

Query Match      32.2%; Score 155; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 6.3e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 DPVDVAVINTTCNYGQVVAAALNATDPCGAAQ 44
Db 1 DPVDVAVINTTCNYGQVVAAALNATDPCGAAQ 30

Search completed: June 22, 2004, 17:27:37
Job time : 4.42692 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 8.89146 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-22

Perfect score: 481

Sequence: 1 VAMSLTVGAGVASADPVDVAV.....PGAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	481	100.0	96	12	US-09-886-349A-22
2	481	100.0	96	14	US-10-193-002-97
3	481	100.0	96	14	US-10-084-843-102
4	481	100.0	96	14	US-10-098-732A-22
5	481	100.0	110	9	US-09-791-171-151
6	481	100.0	110	9	US-09-804-980-151
7	481	100.0	110	12	US-09-943-443-151
8	424	88.1	92	14	US-10-098-732A-63
9	423	87.9	813	15	US-10-369-983-15
10	419.5	87.2	331	9	US-09-287-849-4
11	419.5	87.2	331	14	US-10-359-460-4
12	419.5	87.2	331	15	US-10-359-459-4
13	419.5	87.2	368	9	US-09-287-849-24
14	419.5	87.2	368	14	US-10-359-460-24
15	419.5	87.2	433	9	US-09-287-849-14

16	419.5	87.2	433	14	US-10-359-460-14	Sequence 14, Appl
17	419.5	87.2	856	9	US-09-287-849-12	Sequence 12, Appl
18	419.5	87.2	856	14	US-10-359-460-12	Sequence 12, Appl
19	419	87.1	82	12	US-09-872-186-7	Sequence 7, Appl
20	419	87.1	87	14	US-10-098-732A-61	Sequence 61, Appl
21	419	87.1	282	14	US-10-098-732A-77	Sequence 77, Appl
22	419	87.1	283	14	US-10-098-732A-57	Sequence 57, Appl
23	419	87.1	290	14	US-10-098-732A-51	Sequence 51, Appl
24	419	87.1	299	9	US-09-287-849-19	Sequence 19, Appl
25	419	87.1	299	12	US-09-886-349A-47	Sequence 47, Appl
26	419	87.1	299	14	US-10-359-460-19	Sequence 19, Appl
27	419	87.1	299	14	US-10-098-732A-47	Sequence 47, Appl
28	419	87.1	600	9	US-09-287-849-22	Sequence 22, Appl
29	419	87.1	600	14	US-10-359-460-22	Sequence 22, Appl
30	419	87.1	710	9	US-09-287-849-16	Sequence 16, Appl
31	419	87.1	710	12	US-09-886-349A-49	Sequence 49, Appl
32	419	87.1	710	14	US-10-359-460-16	Sequence 16, Appl
33	419	87.1	710	14	US-10-098-732A-49	Sequence 49, Appl
34	155	32.2	30	13	US-10-005-920-5	Sequence 5, Appl
35	132	27.4	26	13	US-10-005-920-4	Sequence 4, Appl
36	125	26.0	24	13	US-10-005-920-1	Sequence 1, Appl
37	121	25.2	24	13	US-10-005-920-2	Sequence 2, Appl
38	107	22.2	22	13	US-10-005-920-3	Sequence 3, Appl
39	104	21.6	20	14	US-10-193-002-115	Sequence 115, App
40	104	21.6	20	14	US-10-084-843-120	Sequence 120, App
41	93	19.3	20	14	US-10-193-002-54	Sequence 54, Appl
42	93	19.3	20	14	US-10-084-843-54	Sequence 54, Appl
43	75	15.6	296	14	US-10-128-714-3095	Sequence 3095, Ap
44	75	15.6	296	14	US-10-128-714-8095	Sequence 8095, Ap
45	75	15.6	324	12	US-10-425-114-47182	Sequence 47182, A

ALIGNMENTS

RESULT 1
US-09-886-349A-22
; Sequence 22, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB8.4 (DPV)
US-09-886-349A-22

Query Match	100.0%	Score 481;	DB 12;	Length 96;
Best Local Similarity	100.0%	Pred. No. 3.8e-45;		
Matches 96;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	VAMSLTVGAGVASADPVDVAVINTTCNYGVVVAALNATDPCGAAQFNASPVAQSVLRFLA	60	
Db	1	VAMSLTVGAGVASADPVDVAVINTTCNYGVVVAALNATDPCGAAQFNASPVAQSVLRFLA	60	
Qy	61	APPQRAAMAQAQAVEGAAQYIGLVESVAGSCNNY	96	
Db	61	APPQRAAMAQAQAVEGAAQYIGLVESVAGSCNNY	96	

```

RESULT 2
US-10-193-002-97
; Sequence 97, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 97:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 97:
US-10-193-002-97
Query Match 100.0%; Score 481; DB 14; Length 96;
Best Local Similarity 100.0%; Pred No. 3.8e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLRNFLA 60
Db 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLRNFLA 60

QY 61 APPQRAAMAALQAVPGAQYIGLVESVAGSCNNY 96
Db 61 APPQRAAMAALQAVPGAQYIGLVESVAGSCNNY 96

RESULT 3
US-10-084-843-102
; Sequence 102, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

```

```

; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-10-084-843-102
Query Match 100.0%; Score 481; DB 14; Length 96;
Best Local Similarity 100.0%; Pred. No. 3.8e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLRNFLA 60
Db 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLRNFLA 60

QY 61 APPQRAAMAALQAVPGAQYIGLVESVAGSCNNY 96
Db 61 APPQRAAMAALQAVPGAQYIGLVESVAGSCNNY 96

RESULT 4
US-10-098-732A-22
; Sequence 22, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29

```



```

; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB8.4 (DPV)
US-10-098-732A-22

Query Match      100.0%; Score 481; DB 14; Length 96;
Best Local Similarity 100.0%; Pred. No. 3.8e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNAPSPVAQSYLRNFLA 60
Db 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNAPSPVAQSYLRNFLA 60

QY 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
Db 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96

RESULT 5
US-09-791-171-151
; Sequence 151, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 151
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-151

Query Match      100.0%; Score 481; DB 9; Length 110;
Best Local Similarity 100.0%; Pred. No. 4.5e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNAPSPVAQSYLRNFLA 60
Db 15 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNAPSPVAQSYLRNFLA 74

QY 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
Db 75 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 110
```

```

RESULT 6
US-09-804-980-151
; Sequence 151, Application US/09804980
; Publication No. US20030147897A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; APPLICANT: Anderson, Peter
; TITLE OF INVENTION: M. Tuberculosis Antigens
; FILE REFERENCE: 670001-2002.4
; CURRENT APPLICATION NUMBER: US/09/804,980
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 151
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-804-980-151

Query Match      100.0%; Score 481; DB 10; Length 110;
Best Local Similarity 100.0%; Pred. No. 4.5e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNAPSPVAQSYLRNFLA 60
Db 15 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNAPSPVAQSYLRNFLA 74

QY 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
Db 75 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 110

RESULT 7
US-09-943-443-151
; Sequence 151, Application US/09943443
; Publication No. US20030017172A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/943,443
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 151
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-943-443-151

Query Match      100.0%; Score 481; DB 12; Length 110;
Best Local Similarity 100.0%; Pred. No. 4.5e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNAPSPVAQSYLRNFLA 60
Db 15 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNAPSPVAQSYLRNFLA 74
```



```

; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-10-359-460-4

Query Match      87.2%; Score 419.5; DB 14; Length 331;
Best Local Similarity 80.4%; Pred. No. 9.5e-38;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY      5 LTVGAGVASA-----DPVDVINTTCNYGQVVAALNATDPGAAAFNASP 49
|||
Db      129 LTVSVAVSEGKPTKHIQIRSTNKLDPDVAINTTCNYGQVVAALNATDPGAAAFNASP 188
|||

QY      50 VAQSYLRNFLAAPPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 96
|||||
Db      189 VAQSYLRNFLAAPPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 235
|||||

RESULT 12
US-10-359-459-4
; Sequence 4, Application US/10359459
; Publication No. US20040013677A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/10/359,459
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-10-359-459-4

Query Match      87.2%; Score 419.5; DB 15; Length 331;
Best Local Similarity 80.4%; Pred. No. 9.5e-38;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY      5 LTVGAGVASA-----DPVDVINTTCNYGQVVAALNATDPGAAAFNASP 49
|||

```

```

Db      129 LTVSVAVSEGKPTKHIQIRSTNKLDPDVAINTTCNYGQVVAALNATDPGAAAFNASP 188
|||
QY      50 VAQSYLRNFLAAPPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 96
|||||
Db      189 VAQSYLRNFLAAPPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 235
|||||

RESULT 13
US-09-287-849-24
; Sequence 24, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-24

Query Match      87.2%; Score 419.5; DB 9; Length 368;
Best Local Similarity 80.4%; Pred. No. 1.1e-37;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY      5 LTVGAGVASA-----DPVDVINTTCNYGQVVAALNATDPGAAAFNASP 49
|||
Db      130 LTVSVAVSEGKPTKHIQIRSTNKLDPDVAINTTCNYGQVVAALNATDPGAAAFNASP 189
|||

QY      50 VAQSYLRNFLAAPPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 96
|||||
Db      190 VAQSYLRNFLAAPPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 236
|||||

RESULT 14
US-10-359-460-24
; Sequence 24, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849

```

;; PRIOR FILING DATE: 1999-04-07
;; PRIOR APPLICATION NUMBER: US 08/818,112
;; PRIOR FILING DATE: 1997-03-13
;; PRIOR APPLICATION NUMBER: US 08/942,578
;; PRIOR FILING DATE: 1997-10-01
;; PRIOR APPLICATION NUMBER: US 09/025,197
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 09/056,556
;; PRIOR FILING DATE: 1998-04-07
;; PRIOR APPLICATION NUMBER: US 09/223,040
;; PRIOR FILING DATE: 1998-12-30
;; NUMBER OF SEQ ID NOS: 46
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 24
;; LENGTH: 368
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-10-359-460-24

Query Match 87.2%; Score 419.5; DB 14; Length 368;
Best Local Similarity 80.4%; Pred. No. 1.1e-37;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVNTTCNYGQVVAALNATDPGAAAFNASP 49
||| :
DB 130 LTVSVAVSEKPTKHKIQIRSTNKLDPVDVAVNTTCNYGQVVAALNATDPGAAAFNASP 189
||| :
QY 50 VAQSYLRNFLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 96
||| :
DB 190 VAQSYLRNFLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 236
||| :

RESULT 15
US-09-287-849-14
; Sequence 14, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-003020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-14

Query Match 87.2%; Score 419.5; DB 9; Length 433;
Best Local Similarity 80.4%; Pred. No. 1.3e-37;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVNTTCNYGQVVAALNATDPGAAAFNASP 49
||| :
DB 130 LTVSVAVSEKPTKHKIQIRSTNKLDPVDVAVNTTCNYGQVVAALNATDPGAAAFNASP 189
||| :
QY 50 VAQSYLRNFLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 96
||| :
DB 190 VAQSYLRNFLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 236
||| :

Search completed: June 22, 2004, 18:08:02
Job time : 9.89146 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 11.7858 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-22

Perfect score: 481

Sequence: 1 VAMSLTVGAGVASADPVDV.....PGAAQYIGLVSVAGSCNNY 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A: Geneseq 29Jan04: *
1: geneseqp1980s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003bs: *
8: geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	481	100.0	96	2 AAW32349	AAW32349 Mycobacte
2	481	100.0	96	2 AAW32417	AAW32417 Mycobacte
3	481	100.0	96	2 AAW64333	AAW64333 Mycobacte
4	481	100.0	96	2 AAW81700	AAW81700 M. tuberc
5	481	100.0	96	2 AAY38987	AAY38987 M. tuberc
6	481	100.0	96	2 AAY39130	AAY39130 M. tuberc
7	481	100.0	96	4 AAU01894	AAU01894 M. tuberc
8	481	100.0	96	5 AAE29711	AAE29711 Mycobacte
9	481	100.0	96	5 AAE17575	AAE17575 Mycobacte
10	481	100.0	110	2 AAW72936	AAW72936 Mycobacte
11	481	100.0	110	2 AAY21956	AAY21956 Amino aci
12	424	88.1	92	5 AAE29730	AAE29730 Leishmani
13	423	87.9	813	7 ADA26367	ADA26367 Mycobacte
14	419.5	87.2	330	5 AAY74589	AAY74589 Antigenic
15	419.5	87.2	331	2 AAY32060	AAY32060 Mycobacte
16	419.5	87.2	368	5 AAY32069	AAY32069 Mycobacte
17	419.5	87.2	368	5 AAW74598	AAW74598 Antigenic
18	419.5	87.2	433	2 AAY32065	AAY32065 Mycobacte
19	419.5	87.2	433	5 AAU74594	AAU74594 Antigenic
20	419.5	87.2	856	2 AAY32064	AAY32064 Mycobacte
21	419.5	87.2	859	5 AAU74593	AAU74593 Antigenic
22	419	87.1	82	3 AAU96924	AAU96924 M. tuberc
23	419	87.1	87	5 AAE29729	AAE29729 Leishmani
24	419	87.1	282	5 AAE29736	AAE29736 Mycobacte
25	419	87.1	283	5 AAE29727	AAE29727 Leishmani

26 419 87.1 290 5 AAE29725 Aae29725 Leishmani
27 419 87.1 299 2 AAY32067 Ray32067 Mycobacte
28 419 87.1 299 5 AAE29723 Aae29723 Mycobacte
29 419 87.1 299 5 AAE17587 Aae17587 Mycobacte
30 419 87.1 306 5 AAU74596 Aau74596 Antigenic
31 419 87.1 600 2 AAY32068 Ray32068 Mycobacte
32 419 87.1 600 5 AAU74597 Aau74597 Antigenic
33 419 87.1 710 2 AAY32066 Ray32066 Mycobacte
34 419 87.1 710 5 AAE29724 Aae29724 Mycobacte
35 419 87.1 710 5 AAE17588 Aae17588 Mycobacte
36 419 87.1 723 5 AAU74595 Aau74595 Antigenic
37 326 67.8 110 6 AAE37873 Aae37873 Mycobacte
38 229 47.6 60 4 AAB52437 Aab52437 Mycobacte
39 155 32.2 30 5 ABB79492 Abb79492 Mycobacte
40 132 27.4 26 5 ABB79491 Abb79491 Mycobacte
41 125 26.0 24 5 ABB79488 Abb79488 Mycobacte
42 121 25.2 24 5 ABB79489 Abb79489 Mycobacte
43 107 22.2 22 5 ABB79490 Abb79490 Mycobacte
44 104 21.6 20 2 AAW32325 Aaw32325 Mycobacte
45 104 21.6 20 2 AAW32394 Aaw32394 Mycobacte

ALIGNMENTS

RESULT 1
AAW32349
ID AAW32349 standard; protein; 96 AA.
XX AC AAW32349;
XX AC AAW32349;
DT 13-JAN-1998 (first entry)
XX XX
DE Mycobacterium tuberculosis antigen DPV.
XX XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX XX
OS Mycobacterium tuberculosis.
XX XX
FN WO9709429-A2.
XX XX
PD 13-MAR-1997.
XX XX
PF 30-AUG-1996; 96WO-US014675.
XX XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-MAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX XX
PA (CORI-) CORIXA CORP.
XX XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX XX
DR WPI; 1997-192904/17.
DR N-PSDB; AAT91424.
XX XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
PT useful for diagnosis of M. tuberculosis infection.
XX XX
PS Example 1; Page 146; 190pp; English.
XX XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,
CC DPV. The immunogenic polypeptide can be used to diagnose M.tuberculosis
CC infection by forming complexes with specific antibodies in the sample.
CC Fragments of DNA encoding the immunogenic polypeptide can be used as
CC diagnostic primers or probes and agents that bind to the antigen,

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

SQ Sequence 96 AA;

Query Match 100.0%; Score 481; DB 2; Length 96;
 Best Local Similarity 100.0%; Pred. NO. 8.6e-47;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60
 |||||
 DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60
 |||||

QY 61 APPPQRAAQAQIQAVPGAAQYIGLVESVAGSCNNY 96
 |||||
 DB 61 APPPQRAAQAQIQAVPGAAQYIGLVESVAGSCNNY 96
 |||||

RESULT 2
 AAW32417
 ID AAW32417 standard; protein; 96 AA.

XX AAW32417;

XX 06-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen DPV.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX OS

XX W09709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US014674.

XX 01-SEP-1995; 95US-00523436.

XX 22-SEP-1995; 95US-00533634.

XX 22-MAR-1996; 96US-00620874.

XX 05-JUN-1996; 96US-00659683.

XX 12-JUL-1996; 96US-00680574.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky VA, Dillon DC, Campos-Neto A, Houghton R;

PI Vedvick TH, Twardzik DR;

XX WPI; 1997-192903/17.

XX N-PSDB; AAT91488.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.

XX Example 1; Page 134; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its

CC variant differing only in conservative substitutions and/or

CC modifications). The present sequence represents a M.tuberculosis antigen,

CC DPV. The immunogenic protein, and fusion proteins containing one or more

Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60
 |||||
 DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60
 |||||

QY 61 APPPQRAAQAQIQAVPGAAQYIGLVESVAGSCNNY 96
 |||||
 DB 61 APPPQRAAQAQIQAVPGAAQYIGLVESVAGSCNNY 96
 |||||

RESULT 3
 AAW64333
 ID AAW64333 standard; protein; 96 AA.

XX AAW64333;

XX 17-OCT-2003 (revised)

DT 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen DPV.

XX Tuberculosis; infection; diagnosis; antigen; DPV.

XX Mycobacterium tuberculosis; strain H37Rv.

XX W09816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018214.

XX 11-OCT-1996; 96US-00729622.

XX 13-MAR-1997; 97US-00818111.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PI Vedvick TS, Twardzik DR, Lodes MJ;

XX WPI; 1998-251292/22.

XX N-PSDB; AAV44392.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to

PT develop products for the detection of M. tuberculosis infection and

PT diagnosis of tuberculosis.

XX Example 1; Page 129-130; 250pp; English.

XX This polypeptide comprises the Mycobacterium tuberculosis antigen DPV. A
 CC DNA sequence (see AAV44392) coding for DPV was isolated from a M.
 CC tuberculosis genomic library using a probe based on an isolated N-
 CC terminal peptide (see AAW64341). The invention relates to compositions
 CC and methods for diagnosing tuberculosis. It provides polypeptides (see
 CC AAW64291-W64379) comprising an antigenic portion of a soluble M.
 CC tuberculosis antigen, or an immunogenic portion of an M. tuberculosis
 CC antigen, as well as DNA sequences encoding such polypeptides, recombinant
 CC expression vectors and transformed or transfected host cells. Also
 CC claimed are methods and diagnostic kits for detecting M. tuberculosis
 CC infection in a patient using these polypeptides, antibodies or
 CC oligonucleotide probes and primers, for the diagnosis of tuberculosis.
 CC (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 96 AA;

Query Match 100.0%; Score 481; DB 2; Length 96;

Best Local Similarity 100.0%; Pred. NO. 8.6e-47;

Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60
 |||||

DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60
 |||||

Query Match 100.0%; Score 481; DB 2; Length 96;
 Best Local Similarity 100.0%; Pred. NO. 8.6e-47;


```

PR 18-FEB-1998; 98US-00025197.
PR 05-MAY-1998; 98US-00072967.
XX PA
XX (CORI-) CORIXA CORP.
XX
XX Read SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX
XX WPI; 1998-527409/44.
DR N-PSDB; AAZ19302.
XX
XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
PT tests and protective or therapeutic vaccines or compositions.
XX
XX Example 1; Page 119; 299pp; English.
XX
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
CC polypeptides fragments, can be used in pharmaceutical compositions or
CC vaccines to generate a protective or therapeutic immune response to M.
CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
CC killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
CC to AAZ19460 and AAZ39083 to AAZ39225 are used in the exemplification of
CC the present invention
XX
XX Sequence 96 AA;
XX
Query Match 100.0%; Score 481; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.6e-47;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPCGAAQFNASPVQAQSYLRNFLA 60
Db 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPCGAAQFNASPVQAQSYLRNFLA 60
QY 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
Db 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
XX
RESULT 7
AAU01894
ID AAU01894 standard; protein; 96 AA.
XX
XX AAU01894;
XX
XX 29-AUG-2001 (first entry)
XX
XX M. tuberculosis partial antigen Mtb8.4 (DPV).
XX
XX Mtb8.4; DPV; antigen; vaccine; tuberculosis; AIDS;
XX acquired immunodeficiency disease.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200124820-A1.
XX
XX 12-APR-2001.
XX
XX 10-OCT-2000; 2000WO-US028095.
XX
XX 07-OCT-1999; 99US-0158338P.
XX
XX 07-OCT-1999; 99US-0158425P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
XX WPI; 2001-290576/30.
XX
XX N-PSDB; AAS03785.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens.
XX
XX Disclosure; Page 155; 168pp; English.
XX
XX The sequence represents Mycobacterium tuberculosis Mtb8.4 (also known as
CC DPV), an M. tuberculosis antigen. Compositions comprising at least 2
CC heterologous antigens, as a fusion protein, and vectors expressing the
CC fusion proteins are used as vaccines to prophylactically immunise mammals
CC (especially humans) against infection by Mycobacteria. The compositions
CC contain at least 2 heterologous antigens that increase the serological
CC sensitivity of individuals infected with tuberculosis, a disease
CC frequently affecting patients with acquired immunodeficiency disease,
CC AIDS
XX
XX Sequence 96 AA;
XX
Query Match 100.0%; Score 481; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.6e-47;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPCGAAQFNASPVQAQSYLRNFLA 60
Db 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPCGAAQFNASPVQAQSYLRNFLA 60
QY 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
Db 61 APPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
XX
RESULT 8
AAE29711
ID AAE29711 standard; protein; 96 AA.
XX
XX AAE29711;
XX
XX 27-JAN-2003 (first entry)
XX
XX Mycobacterium tuberculosis DPV antigenic protein.
XX
XX Vaccine; immunity; diagnostic agent; gene therapy; DPV antigen; Mtb8.4.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200272792-A2.
XX
XX 19-SEP-2002.
XX
XX 13-MAR-2002; 2002WO-US008223.
XX
XX 13-MAR-2001; 2001US-0275837P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Brannon M, Guderian J;
XX WPI; 2002-759844/82.
XX
XX N-PSDB; AAD47087.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX tuberculosis.
XX
XX Disclosure; Page 100; 155pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a
XX fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX polynucleotide sequence encoding an antigen or an antigenic fragment from
XX Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX polypeptide or its fragment. The Leishmania polynucleotide is selected
XX from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
XX

```


CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is M.
CC tuberculosis DPV antigenic protein. DPV is also referred to as MTB8.4
XX
XX
SQ Sequence 96 AA;

Query Match 100.0%; Score 481; DB 5; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.6e-47;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGAFNASEPVAQSYLRNFLA 60
DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGAFNASEPVAQSYLRNFLA 60
QY 61 APPQRAAMAQLOQAVPGAQYIGLVESVAGSCNNY 96
DB 61 APPQRAAMAQLOQAVPGAQYIGLVESVAGSCNNY 96

RESULT 9
AAE17575
ID AAE17575 standard; protein; 96 AA.
XX
AC AAE17575;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species MTB8.4 (DPV) protein.
XX
XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB8.4; DPV protein.
KW
XX
OS Mycobacterium sp.
XX
XX WO200198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US019959.
XX
XX 20-JUN-2000; 2000US-00597796.
PR
XX 01-FEB-2001; 2001US-0265737P.
XX
PA (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Alderson M;
PI
XX WPI; 2002-147798/19.
DR
XX N-PSDB; AAD28345.
XX

Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a subject.
XX
XX
PS Claim 9; Page 116; 136pp; English.
XX
XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the

CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
XX MTB8.4 (DPV) protein
XX
SQ Sequence 96 AA;

Query Match 100.0%; Score 481; DB 5; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.6e-47;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGAFNASEPVAQSYLRNFLA 60
DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGAAGAFNASEPVAQSYLRNFLA 60
QY 61 APPQRAAMAQLOQAVPGAQYIGLVESVAGSCNNY 96
DB 61 APPQRAAMAQLOQAVPGAQYIGLVESVAGSCNNY 96

RESULT 10
AAW72936
ID AAW72936 standard; protein; 110 AA.
XX
AC AAW72936;
XX
DT 21-JAN-1999 (first entry)
XX
DE Mycobacterium tuberculosis antigen CFP98.
XX
XX Mycobacterium tuberculosis; antigen; vaccine; immunological; immunogen;
KW infection.
KW
XX Mycobacterium tuberculosis.
OS
XX WO9844119-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-DK000132.
XX
XX 02-APR-1997; 97DK-00000376.
PR
XX 18-APR-1997; 97US-0044624P.
PR
XX 10-NOV-1997; 97DK-00001277.
PR
XX 05-JAN-1998; 98US-0070488P.
XX
PA (STAT-) STATENS SERUM INST.
XX
XX Andersen P, Nielsen R, Rosenkrands I, Weldingh K, Rasmussen PB;
PI Oettinger T, Florio W;
XX
XX WPI; 1998-542705/46.
DR
XX N-PSDB; AAV63946.
XX

New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis.
XX
XX Claim 1; Page 225; 163pp; English.
XX
XX The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis, M.
CC africanum or M. bovis
XX
XX Sequence 110 AA;

Query Match 100.0%; Score 481; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. NO. 1e-46;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCAAQFNASPVAQSYLRNFLA 60
 |||||
 DB 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCAAQFNASPVAQSYLRNFLA 74
 |||||

QY 61 APPQRAAAQLOAVFGAAQYIGLVESVAGSCNNY 96
 |||||
 DB 75 APPQRAAAQLOAVFGAAQYIGLVESVAGSCNNY 110
 |||||

RESULT 11
 AAY21956
 ID AAY21956 standard; protein; 110 AA.
 XX
 AC AAY21956;

XX
 XX
 XX 06-SEP-1999 (first entry)

XX
 XX Amino acid sequence of antigen CFP83.

XX Immunogenic; Mycobacterium tuberculosis; immune response; infection;
 KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;
 KW pharmaceutical; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;
 KW CFP7B; CFP19; CFP27; CFP30A; RDI-ORF; CFP10A; CFP16; CFP19; CFP23;
 KW CFP25A; CFP30B; CFP7B.

XX Mycobacterium tuberculosis.

XX
 XX WO9924577-A1.

XX 20-MAY-1999.

XX 08-OCT-1998; 98WO-DK000438.

XX 10-NOV-1997; 97DK-00001277.

XX 05-JAN-1998; 98US-0070488P.

XX 01-APR-1998; 98WO-DK000132.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Skjot R;

XX WPI; 1999-347282/29.

XX N-PSDB; AAX81095.

XX New immunogenic fragment of Mycobacterium tuberculosis.

XX Example 3; Page 242; 265pp; English.

XX The invention describes a substantially pure immunogenic polypeptide
 CC fragment (I) from Mycobacterium tuberculosis that is able to evoke a
 CC protective immune response against infections by mycobacteria belonging
 CC to the tuberculosis complex. The invention provides a (1) fusion
 CC polypeptide comprising at least one polypeptide fragment (I) and at least
 CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell
 CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second
 CC different amino acid sequence from M. tuberculosis, and/or including a
 CC sequence which protects the first amino acid sequence from in vivo
 CC degradation or post-translational processing; (3) a nucleic acid fragment
 CC that encodes the above polypeptides. The polypeptides and nucleic acid
 CC are useful as pharmaceuticals, for diagnosis of and as antigens for
 CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or
 CC bovis. The polypeptides are also useful for diagnosing ongoing or
 CC previous sensitization in an animal with bacteria belonging to the
 CC tuberculosis complex. The invention also describes the use of CFP7A or
 CC CFP30A or a T-cell epitope of for the induction of a strong immune
 CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell
 CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin
 CC test; use of CFP27, CFP30A, RDI-ORF2, RDI-ORF3, RDI-ORF5, MPT59-ESAT6,
 CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-

CC cell epitope of for the preparation of an immunological composition; and
 CC for the preparation of a subunit vaccine

XX Sequence 110 AA;

Query Match 100.0%; Score 481; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. NO. 1e-46;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCAAQFNASPVAQSYLRNFLA 60
 |||||
 DB 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCAAQFNASPVAQSYLRNFLA 74
 |||||

QY 61 APPQRAAAQLOAVFGAAQYIGLVESVAGSCNNY 96
 |||||
 DB 75 APPQRAAAQLOAVFGAAQYIGLVESVAGSCNNY 110
 |||||

RESULT 12
 AAE29730
 ID AAE29730 standard; protein; 92 AA.
 XX
 AC AAE29730;

XX 29-AUG-2003 (revised)
 DT 27-JAN-2003 (first entry)

XX Leishmania sp. MAPS (N10) -Mycobacterium sp. DPV-AC fusion protein.

XX Vaccine; immunity; diagnostic agent; gene therapy; MAPS; DPV;
 KW fusion protein.

XX Mycobacterium sp.

XX Leishmania sp.

XX Chimeric.

XX WO200272792-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US008223.

XX 13-MAR-2001; 2001US-0275837P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Brannon M, Guderian J;

XX WPI; 2002-759844/82.

XX N-PSDB; AAD47109.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif.

XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity

XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium

XX tuberculosis.

XX Example 4; Page 127-128; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC Leishmania species thiol specific antioxidant [TSA; MAPS (N10)] -
 CC Mycobacterium sp. DPV-AC (codon optimised) fusion protein. This sequence
 CC comprises the first ten amino acids of Leishmania species TSA (MAPS) at

Db 129 LTVSVAVSEKPTKHKIIRSTNKLDPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 188
QY 50 VAQSYLRNFLAAPPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 96
Db 189 VAQSYLRNFLAAPPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 235

RESULT 15

AAV32060

ID AAV32060 standard; protein; 331 AA.

XX AC AAV32060;

XX DT 17-JAN-2000 (first entry)

XX DE Mycobacterium tuberculosis antigen fusion protein Mtb39A.

XX KW Tuberculosis; antigen; fusion protein; Mtb39A; ERD14; DPV; MTI;

XX KW diagnosis; therapy; vaccine; immunogen.

XX OS Mycobacterium tuberculosis.

XX FH Key

XX FT Peptide

XX FT 1. .7

XX FT /note= "Met/His tag"

XX FT Protein

XX FT 8. .151

XX FT /note= "ERD14"

XX FT Protein

XX FT 154. .235

XX FT /note= "DPV"

XX FT Protein

XX FT 238. .331

XX FT /note= "MTI"

XX PN WO9951748-A2.

XX PD 14-OCT-1999.

XX PF 07-APR-1999; 99WO-US007717.

XX PR 07-APR-1998; 98US-00056556.

XX PR 30-DEC-1998; 98US-00223040.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky YAW, Alderson M, Campos-Neto A;

XX DR WPI; 1999-601610/51.

XX DR N-PSDB; AAZ20195.

XX PT New fusion proteins useful for diagnosis, prevention and treatment of

XX PT tuberculosis.

XX PS Claim 1; Fig 2; 83pp; English.

XX CC This sequence represents a recombinant Mycobacterium tuberculosis tri-

XX CC antigen fusion protein, termed Mtb39A, composed of the antigens ERD14,

XX CC DPV and MTI. The fusion protein is expressed in host cells using a vector

XX CC carrying a polynucleotide (see AAZ20195) comprising the 3 coding

XX CC sequences for the antigens. The invention provides fusion proteins (see

XX CC AAY32059-71) containing at least 2 M. tuberculosis antigens. The new

XX CC fusion proteins and polynucleotides encoding them are useful as vaccines

XX CC for preventing tuberculosis (claimed), for diagnosis (via in vitro assays

XX CC or intradermal skin tests for detection of anti-M. tuberculosis

XX CC antibodies), monitoring of disease progression, and treatment of

XX CC tuberculosis. They are more effective immunogens than mixtures of the

XX CC individual protein components

XX SQ Sequence 331 AA;

XX Query Match 87.2%; Score 419.5; DB 2; Length 331;

XX Best Local Similarity 80.4%; Pred. No. 3.8e-39;

XX Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

XX 5 LTVGAGVASA-----DPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 49

Db 129 LTVSVAVSEKPTKHKIIRSTNKLDPVDVAVINTTCNYGQVVAALNATDPGAAAFNASP 188
QY 50 VAQSYLRNFLAAPPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 96
Db 189 VAQSYLRNFLAAPPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 235

Search completed: June 22, 2004, 17:15:45
Job time : 12.7858 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 2.68596 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-22

Perfect score: 481

Sequence: 1 VAMSLTVGAGVASADPVDVAV.....PGAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	481	100.0	110	2	F70875	hypothetical prote
2	73.5	15.3	1477	2	AG3009	polyketide synthet
3	73.5	15.3	1489	2	G98274	hypothetical prote
4	73	15.2	626	2	S36364	glucan 1,4-alpha-g
5	73	15.2	626	2	T49625	glucan 1,4-alpha-g
6	72.5	15.1	197	2	A86430	F26G16.12 protein
7	72	15.0	196	2	E81966	probable periplasm
8	72	15.0	344	2	A70771	probable hydrolase
9	72	15.0	345	1	VABPA7	major capsid prote
10	72	15.0	398	1	VBSPA7	minor capsid prote
11	70.5	14.7	462	2	B83123	probable endoprote
12	70.5	14.7	1832	2	T31113	mucin-like glycopr
13	70	14.6	196	2	D81022	conserved hypothet
14	70	14.6	451	2	D88395	protein F53A3.6 li
15	69.5	14.4	288	2	G84351	thiamin monophosph
16	69	14.3	227	2	G70555	hypothetical prote
17	69	14.3	244	2	D96707	probable zinc fing
18	69	14.3	343	1	D33858	mercury(II) reduct
19	69	14.3	655	2	AD1598	DNA gyrase-like pr
20	69	14.3	655	2	AF1235	DNA gyrase-like pr
21	68.5	14.2	362	2	G87470	Spherulin family secret
22	68	14.1	243	2	B29624	shperulin lb precu
23	68	14.1	347	2	T46192	acetylglutamate ki
24	67.5	14.0	326	2	AQ3269	malate dehydrogena
25	67.5	14.0	865	2	A47282	calcium-binding pr
26	67.5	14.0	1400	2	T31555	hypothetical prote
27	67	13.9	465	2	G02738	FRBAC-4 - human
28	67	13.9	1607	2	G87259	conserved hypothet
29	67	13.9	1607	2	T02837	long chain fatty a

30 66.5 13.8 314 2 T03554 potential transcri
31 66.5 13.8 353 2 B84393 D-lactate dehydrog
32 66.5 13.8 399 2 S53975 probable membrane
33 66.5 13.8 732 2 A40839 ornithine decarbox
34 66.5 13.8 842 2 T36379 glycoprotein H pre
35 66 13.7 228 2 T36379 probable peptidase
36 66 13.7 351 2 C83089 histidinol-phospha
37 66 13.7 1226 2 T24045 hypothetical prote
38 66 13.7 1226 2 T24045 hypothetical prote
39 66 13.7 3013 2 AB0480 probable cation tr
40 66 13.7 3184 1 WMBEH6 UL36 protein - hum
41 65.5 13.6 186 2 T03570 hypothetical prote
42 65.5 13.6 287 2 G86728 alpha-subuni L-ser
43 65.5 13.6 484 2 A83082 probable porin PA4
44 65.5 13.6 663 2 S55164 hypothetical prote
45 65 13.5 332 2 A75352 hypothetical prote

ALIGNMENTS

RESULT 1

F70875

hypothetical protein Rv1174c - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C/Accession: F70875

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: F70875

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-110 <COL>

A/Cross-references: GB:AL010186; GB:AL123456; NID:g3261493; PIDN:CAA15851.1; PID:g2595955

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: Rv1174c

C/Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1174c

Query Match 100.0%; Score 481; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 28-41;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCGAAQFNASPVQSVLRNFLA 60
Db 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCGAAQFNASPVQSVLRNFLA 74

Qy 61 APPPQRAAMAAQVAPGAAQYIGLVESVAGSCNNY 96

Db 75 APPPQRAAMAAQVAPGAAQYIGLVESVAGSCNNY 110

RESULT 2

AG3009

polyketide synthetase Atu3681 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: AG3009

R/Wood, D.W.; Serubai, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan; Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.

ster, E.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Accession: AG3009

A/Status: preliminary

C;Accession: A86430
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ausen, N.F.; Hughes, B.; Huijzer, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Cazzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
R;Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86430
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-197 <STO>
A;Cross-references: GB:AE005172; NID:G6634774; PIDN:AAE19754.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 15.1%; Score 72.5; DB 2; Length 197;
Best Local Similarity 31.0%; Pred. No. 4.3;
Matches 26; Conservative 8; Mismatches 23; Indels 27; Gaps 5;

QY 26 NYGQVVAALNATDPG-----AAQNPASPVQSVLRNFLAAPPQ-----RAAQAQLQ 74
DB 41 NYESIIVSLVSDPGTINSMPGQY---PYDPYIRS-IFAPPQPYTGVLHQLMLGVQQQ 96
QY 75 AVP-----GAAQYIGLV 86
DB 97 GVPLPSDAVEPFVFNKQYHGIL 120

RESULT 7
E81966
probable periplasmic transport protein NMA0488 [imported] - Neisseria meningitidis (stra
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: E81966
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: E81966
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <PAR>
A;Cross-references: GB:AL162753; GB:AL157959; NID:G7379120; PIDN:CAB83781.1; PID:G737923
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA0488

Query Match 15.0%; Score 72; DB 2; Length 196;
Best Local Similarity 32.6%; Pred. No. 4.8;
Matches 28; Conservative 11; Mismatches 43; Indels 4; Gaps 2;

QY 10 GVASADPVDVINTCNVGVVAALNATDPGAA---AQNPASPVQSVLRNFLAAPPQ 66
DB 18 GMAFAAPADAVNQIRQNAQTQVLSILKSGDANTARQAEAYAIYPDFQFMTALAVGNPW 77
QY 67 AAMAAQLQAVPGAAQVIGLVESVAGS 92
DB 78 TASDAQKQALAKEFQTL-LIRYSQT 102

RESULT 8
A70771
probable hydrolase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: A70771
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.C.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:96343230
A;Accession: A70771
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-344 <COL>
A;Cross-references: GB:Z73902; GB:AL123456; NID:G3261576; PIDN:CAA98097.1; PID:G245027; I
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv1333

Query Match 15.0%; Score 72; DB 2; Length 344;
Best Local Similarity 34.1%; Pred. No. 8.7;
Matches 29; Conservative 7; Mismatches 29; Indels 20; Gaps 3;

QY 2 AMSLTVGAGVASAD-----PVDVINTCNVGVVAALNATDPGAAQFNASPVQSV 54
DB 155 AGALKGGVGTASATLQSGVTGVGLAVN-----AAGNVVDPATGLPMMADLVGEFA 205
QY 55 LRNFLAAPPQRAAQAQLQAVPGA 79
DB 206 LR-----APPAEQIALAQJSSPLGA 226

RESULT 9
VABPA7
major capsid protein 10A - phase T7
C;Species: phage T7
A;Note: host Escherichia coli
C;Date: 13-Jun-1983 #sequence_revision 30-Sep-1990 #text_change 23-Jul-1999
C;Accession: A04344; S42325
R;Dunn, J.J.; Thompson, K.
submitted to the Nucleic Acid Sequence Database, September 1982
A;Reference number: A94615
A;Accession: A04344
A;Molecule type: DNA
A;Residues: 1-345 <DUN>
R;Dunn, J.J.; Studier, F.W.
J. Mol. Biol. 166, 477-535, 1983
A;Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 ge
A;Reference number: S42283; MUID:83241725; PMID:6864790
A;Accession: S42325
A;Molecule type: DNA
A;Residues: 1-345 <DUW>
A;Cross-references: EMBL:V01146; NID:G431187; PIDN:CAA24427.1; PID:G15604
A;Note: the authors did not translate the codon for residue 1
C;Genetics:
A;Gene: 10A
A;Map position: 57.51-60.49
C;Superfamily: phage T7 major capsid protein 10A
C;Keywords: capsid protein

Query Match 15.0%; Score 72; DB 1; Length 345;
Best Local Similarity 31.6%; Pred. No. 8.8;
Matches 31; Conservative 8; Mismatches 35; Indels 24; Gaps 5;

QY 8 GAGVASADPVDVINTCNVGVVAALNATDPGAAQFNASPVQSVLRN 57
DB 158 GLGTAT-----VIETQNKAAALTDQVALGKEIIAAL--TKARAALTKNVVPAAD---RV 206
QY 58 FLAAPPQRAAQAQLQAVPGAQYIGLVESVAGSCNN 95
DB 207 FYCDPSYSAILAA---LMPNAAAYALIDPEKGSIRN 241

RESULT 10
VBBPA7
minor capsid protein 10B - phase T7
C;Species: phage T7

A;Note: host Escherichia coli
 C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 23-Jul-1999
 C;Accession: B04344; S42326
 R;Dunn, J.J.; Thompson, K.
 submitted to the Nucleic Acid Sequence Database, September 1982
 A;Reference number: A94615
 A;Accession: B04344
 A;Molecule type: DNA
 A;Residues: 1-398 <DUN>
 R;Dunn, J.J.; Studier, F.W.
 J. Mol. Biol. 166, 477-535, 1983
 A;Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genes
 A;Reference number: S42283; MUID:83241725; PMID:6864790
 A;Accession: S42326
 A;Status: preliminary; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-398 <DU2>
 A;Cross-references: EMBL:V01146; NID:G431187; PIDN:CAA24428.1; PID:G431193
 C;Genetics:
 A;Gene: 108
 A;Map position: 57.51-60.49
 A;Note: translation of the nucleotide sequence involves a -1 frameshift within codon 341
 C;Superfamily: phage T7 major capsid protein 10A
 C;Keywords: capsid protein

Query Match 15.0%; Score 72; DB 1; Length 398;
 Best Local Similarity 31.6%; Pred. No. 10;
 Matches 31; Conservative 8; Mismatches 35; Indels 24; Gaps 5;

QY 8 GAGVAGADPVDVINTTCNYG-----QVVAALNATDPCGAAQFNASPVASQYLRN 57
 Db 158 GLGTAT-----VIETONKAALTQVALGKEIIAAL--TKARAALTKNYVPAAD---RV 206

QY 58 FLAAPPQBAWAQALQAVPGAAQYIGLVESVAGSCNN 95
 Db 207 FYCDPSYSALIAA---LMPNAAANYAALIDPEKGSIRN 241

RESULT 11
 B83123
 Probable endoproteinase Arg-C precursor PA4175 [imported] - Pseudomonas aeruginosa (strain PA01)
 C;Species: Pseudomonas aeruginosa
 C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C;Accession: B83123
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Broman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
 Nature 405, 959-964, 2000
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A;Reference number: B82950; MUID:20437337; PMID:10984043
 A;Accession: B83123
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-462 <STO>
 A;Cross-references: GB:AE004834; GB:AE004091; NID:G9950382; PIDN:AAG07562.1; GSPDB:GN001
 A;Experimental source: strain PA01
 C;Genetics:
 A;Gene: PA4175

Query Match 14.7%; Score 70.5; DB 2; Length 462;
 Best Local Similarity 29.1%; Pred. No. 17;
 Matches 23; Conservative 9; Mismatches 34; Indels 13; Gaps 2;

QY 20 VINTTCNYGVVAALNATD-PCGAAQFNASPVASQYLR-----NFLAAPPPQR 66
 Db 12 VLALAGASQALAPGASMGADGVAVLQASPASTGHAFNPFNAISAAIGHFAAPPARR 71

QY 67 AAMAALQAVPGAAQYIGL 85
 Db 72 VARAPLAPKPTPLQGV 90

RESULT 12
 D88395
 Protein F53A3.6 [imported] - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C;Accession: D88395
 R;anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A;Reference number: A75000; MUID:99069613; PMID:9851916

T31113
 mucin-like glycoprotein 900 - Cryptosporidium parvum
 C;Species: Cryptosporidium parvum
 C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C;Accession: T31113
 R;Barnes, D.A.; Bonnin, A.; Huang, J.X.; Gousset, L.; Wu, J.; Gut, J.; Doyle, P.; Dubremet, M.; Biochem. Parasitol. 96, 93-110, 1998
 A;Title: A novel multi-domain mucin-like glycoprotein of Cryptosporidium parvum mediates adherence
 A;Reference number: Z20989; MUID:99066935; PMID:9851610
 A;Accession: T31113
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1832 <BAR>
 A;Cross-references: EMBL:AF068065; NID:G4063041; PID:G4063042; PIDN:AAC98153.1

Query Match 14.7%; Score 70.5; DB 2; Length 1832;
 Best Local Similarity 28.4%; Pred. No. 74;
 Matches 21; Conservative 12; Mismatches 26; Indels 15; Gaps 2;

QY 31 VAALNATDPCGAAQFNASPVASQY-----LRNFLAA-----PPQRAAAMAALQQA 75
 Db 1690 VSGFATVPTGTAAPKGGVTPESVAAEAARKYFAANVEGEGEGEYPPPPPESSSNTAIOA 1749

QY 76 VPGAQYIGLVESV 89
 Db 1750 AGGASAAVGLVAAV 1763

RESULT 13
 D81022
 conserved hypothetical protein NMB1963 [imported] - Neisseria meningitidis (strain MC58)
 C;Species: Neisseria meningitidis
 C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C;Accession: D81022
 R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Fizzia, M.
 Science 287, 1809-1815, 2000
 A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; Ver
 A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A;Reference number: A81000; MUID:20175755; PMID:10710307
 A;Accession: D81022
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-196 <TET>
 A;Cross-references: GB:AE002544; GB:AE002099; NID:G7227214; PIDN:AAF42292.1; PID:G7227222;
 A;Experimental source: serogroup B, strain MC58
 C;Genetics:
 A;Gene: NMB1963

Query Match 14.8%; Score 70; DB 2; Length 196;
 Best Local Similarity 32.8%; Pred. No. 7.6;
 Matches 28; Conservative 10; Mismatches 44; Indels 4; Gaps 2;

QY 10 GVASADPVDVINTTCNYGVVAALNATDPCGAA---AQFNASPVASQYLRNFLAAPPPQR 66
 Db 18 GMAFAPADAVSIRQATQVLSILKNGDANTARQAEAYAIYFFDFQMTALAVGNPWR 77

QY 67 AAMAALQAVPGAAQYIGLVESVAGS 92
 Db 78 TASDAQKQALAKEFQTL-LIRTSYSGT 102

RESULT 14
 D88395
 Protein F53A3.6 [imported] - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C;Accession: D88395
 R;anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A;Reference number: A75000; MUID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: D88395
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-451 <STO>
A:Cross-references: GB:chr_III; PIDN:AAB70991.1; PID:g2429454; GSPDB:GN00021; CESP:F53A3
C:Genetics:
A:Gene: F53A3.6
A:Map position: 3

Query Match 14.6%; Score 70; DB 2; Length 451;
Best Local Similarity 35.4%; Pred. No. 19;
Matches 29; Conservative 5; Mismatches 34; Indels 14; Gaps 3;
QY 5 LTVGAGVASADPVDVAVITTCNYGQVVAALNATDCAAAQFNASPVASQSYLRNPLAAPP 64
DB 370 LTAGSGSAGIPYFSAL-----SQLNQLGAAAFGAPGTLNGLQFPAN-----AALGP 416
QY 65 CRAAMAAQLQAVPGAQVIGLV 86
DB 417 QLAG-AALLAAVPGAQQQIKVV 437

RESULT 15
G84351
thiamin monophosphate kinase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 07-Jul-2003
C:Accession: G84351
R:NG, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omar, A.D.; Ebhardt, H.; Lowe, T.M.; L
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: G84351
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-288 <STO>
A:Cross-references: GB:AE004437; NID:g10581442; PIDN:ARG20179.1; GSPDB:GN00138
C:Genetics:
A:Gene: thil
C:Superfamily: thiamine monophosphate kinase

Query Match 14.4%; Score 69.5; DB 2; Length 288;
Best Local Similarity 32.0%; Pred. No. 13;
Matches 33; Conservative 15; Mismatches 38; Indels 17; Gaps 6;
QY 1 VAMSLTVGAGVASADPVDVAVITTCNYGQVVAALN---ATDPG-----AAQFNASPV- 50
DB 173 VAAGRVLGAHATAM--MDASDGLARSLHQLAAASDCGMAVDSGRFLFVADALAEVTADE 230
QY 51 -AQSYLRNF--LAAPPQRAAAQQLQAVPGAQVIGLVESVA 90
DB 231 RAVSGGDFELVAAPPPER--VEAFAAVFGSLSVVGRVTTAA 271

Search completed: June 22, 2004, 17:24:58
Job time : 5.68596 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 1.59768 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-22
Perfect score: 481
Sequence: 1 VAMSLITVAGVAGSADPPDAV.....FGAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swiseprot_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	75.5	15.7	621	1 VP40_HSVBC	P54817 bovine herp
2	73.5	15.3	509	1 TSSP_MOUSE	Q9XKE5 mus musculu
3	73	15.2	626	1 AMYG_NEUCR	P14804 neurospora
4	72	15.0	344	1 YD33_MYCTU	Q10644 mycobacteri
5	72	15.0	345	1 VCAA_BPT7	P19726 bacterioph
6	72	15.0	398	1 VCAAB_BPT7	P19727 bacterioph
7	69	14.3	460	1 MEGA_SERMA	P08662 serratia ma
8	68	14.1	183	1 REGA_RHOSU	O82868 rhodorum
9	67.5	14.0	246	1 SRIA_RHYPO	P03350 physarum po
10	67.5	14.0	865	1 CPN_DROME	Q02910 drosophila
11	67	13.9	465	1 FXD1_HUMAN	P16676 homo sapien
12	66.5	13.8	389	1 SCWA_YEAST	Q04951 saccharomyc
13	66.5	13.8	514	1 TSSP_HUMAN	Q9HGE7 homo sapien
14	66.5	13.8	732	1 DCOS_ECOLI	P24169 escherichia
15	66.5	13.8	842	1 VGLH_HSVBC	P27599 bovine herp
16	66	13.7	351	1 H181_PSEAE	Q9NVX0 pseudomonas
17	66	13.7	1609	1 CTPI_MYCLE	O53114 mycobacteri
18	66	13.7	3164	1 TEGU_HSV11	P10220 herpes simp
19	65.5	13.6	663	1 YU09_YEAST	P47005 saccharomyc
20	65	13.5	323	1 MIAA_RALSO	Q8XWB0 ralstonia s
21	65	13.5	618	1 VE1_HPV63	P07847 human papil
22	65	13.5	659	1 AEGA_ECOLI	P37127 escherichia
23	65	13.5	1157	1 SRA4_HUMAN	Q9S104 homo sapien
24	65	13.5	1442	1 PTC1_CHICK	Q90693 gallus gall
25	65	13.5	2038	1 FSH_DROME	P13709 drosophila
26	64.5	13.4	248	1 SRIE_RHYPO	P09351 physarum po
27	64.5	13.4	358	1 LIP_PBEGL	Q05489 pseudomonas
28	64.5	13.4	474	1 NOA1_RAT	Q80W44 rattus norv
29	64.5	13.4	493	1 NOA1_MOUSE	Q9JKN6 mus musculu
30	64.5	13.4	510	1 NOA1_HUMAN	P51513 homo sapien
31	64	13.3	203	1 CUTI_PRRER	Q9Y7G8 pyrenopeziz
32	64	13.3	686	1 SOWD_HUMAN	O95936 homo sapien
33	64	13.3	1434	1 PTC1_MOUSE	Q61115 mus musculu

34	63.5	13.2	430	1 CIWC_HUMAN	Q9HBI5 homo sapien
35	63.5	13.2	430	1 CIWC_RAT	Q9ERL1 rattus norv
36	63	13.1	156	1 I3MS_BRANA	Q00650 brassica na
37	63	13.1	1625	1 CTPI_MYCTU	Q10900 mycobacteri
38	62.5	13.0	3421	1 TEGU_HSVB	P28955 equine herp
39	62	12.9	377	1 NADA_CAUCR	Q9A4C4 caulobacter
40	62	12.9	399	1 ISDF_TREPA	O83525 t ispd/ispf
41	62	12.9	415	1 LMP2_MOUSE	P17047 mus musculu
42	62	12.9	676	1 ICPO_HSVBK	P29836 bovine herp
43	62	12.9	1024	1 TIPT_DROME	Q9U3V5 drosophila
44	62	12.9	1317	1 GAP_CAEEL	P34288 caenorhabdi
45	61.5	12.8	271	1 HXA9_MOUSE	P09631 mus musculu

ALIGNMENTS

RESULT 1
VP40_HSVBC STANDARD; PRT; 621 AA.
AC P54817: O89855;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Capsid protein P40 [Contains: Assemblin (Protease) (EC 3.4.21.97);
DE Capsid assembly protein].
GN UL26.
OS Bovine herpesvirus type 1 (strain Cooper).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=10323;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=96013856; PubMed=7474173;
RA Haanes E.J., Thomsen D.R., Martin S., Homa F.L., Lowery D.E.;
RT "The bovine herpesvirus 1 maturational proteinase and scaffold
proteins can substitute for the homologous herpes simplex virus type
1 proteins in the formation of hybrid type B capsids.";
RL J. Virol. 69:7375-7379(1995).
CC -!- FUNCTION: THE CAPSID ASSEMBLY PROTEIN IS COMPONENT OF THE CAPSID
CORE INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.
CC -!- FUNCTION: ASSEMBLIN IS A PROTEASE WHICH CAN PROTEOLYTICALLY CLEAVE
ITSELF AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.
CC -!- CATALYTIC ACTIVITY: Cleaves -Ala-|-Ser- and -Ala-|-Ala- bonds in
the scaffold protein.
CC -!- ALTERNATIVE PRODUCTS:
Event-Alternative initiation;
Comment=2 isoforms, UL26 (shown here) and UL26.5, are produced
by alternative initiation;
-!- SIMILARITY: Belongs to peptidase family S21.
-!- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sb-sib.ch).

EMBL: U31809; AAA91999.1; -;
EMBL: U31809; AAA92000.1; -;
EMBL: Z78205; CAB01599.1; -;
EMBL: AJ004801; CAA06107.1; -;
HSSP: P16753; 1CMW.
MEROPS: S21.001; -;
InterPro: IPR001847; Peptidase S21.
Pfam: PF00716; Peptidase S21; 1.
PRINTS: PR00236; HSVCAPSIDP40.
KW Capsid assembly; Hydrolase; Serine protease; Coat protein;
Alternative initiation.
CHAIN 1 621 CAPSID PROTEIN P40, ISOFORM UL26.
CHAIN 314 621 CAPSID PROTEIN P40, ISOFORM UL26.5.

FT INIT MET 314 314 FOR ISOFORM UL26.5.
 FT CHAIN 1 254 ASSEMBLIN (PROTEASE).
 FT CHAIN 255 621 CARSID ASSEMBLY PROTEIN.
 FT SITE 254 255 CLEAVAGE (BY THE PROTEASE) (PROBABLE).
 FT ACT SITE 68 68 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 136 136 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 155 155 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 621 AA; AEF6D7C9BD569003 CRC64;
 Query Match 15.7%; Score 75.5; DB 1; Length 621;
 Best Local Similarity 33.0%; Pred. No. 2.6;
 Matches 35; Conservative 6; Mismatches 30; Indels 35; Gaps 5;
 QY 2 AMSLTGAGVASADPV---DAVINTTCYQGVV-----AALNATDPGAAQFNASPVQAQ 52
 Db 313 AMSAAHQAGASAEHPLFAGDYVVVPTAQYNQLVVQARGAMTAAPPAPYFLPAAAAA- 371
 QY 53 SYLRNFLAAPPQ-----RAAAA-----QLQAVPGA 79
 Db 372 -----AAAPPPMGVGAAGAPWHPGYGFPFPPGLESQTMALAGA 411
 RESULT 2
 TSSP_MOUSE STANDARD; PRT; 509 AA.
 ID TSSP_MOUSE STANDARD; PRT; 509 AA.
 AC Q9QXES;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thymus-specific serine protease precursor (EC 3.4.-.-).
 GN PRS16 OR TSSP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryonic thymus;
 RX MEDLINE=20093985; PubMed=10630289;
 RA Carrier A., Nguyen C., Victorero G., Granjeaud S., Rocha D.,
 RA Bernard K., Mizsek A., Ferrier P., Malissen M., Naquet P.,
 RA Malissen B., Jordan B.R.
 RT "Differential gene expression in CD3epsilon- and RAG1-deficient
 RT thymuses: definition of a set of genes potentially involved in
 RT thymocyte maturation."
 RL Immunogenetics 50:255-270(1999).
 CC -!- FUNCTION: Protease that may play a role in T-cell development.
 CC -!- SUBCELLULAR LOCATION: Vesicular, either lysosomal or endosomal.
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in cortical thymic
 CC epithelial cells, with highest expression around vessels and the
 CC thymic capsule.
 CC -!- DEVELOPMENTAL STAGE: Expressed in developing thymus at 14 to 18
 CC dpc, with maximal expression at 16 dpc.
 CC -!- SIMILARITY: Belongs to peptidase family S28.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AJ131775; CAB66137.1;
 DR MEROPS; S28.003;
 DR MGD; MGI:1859181; Prs16.
 DR InterPro; IPR008758; Peptidase_S28.
 DR InterPro; IPR000379; Ser_estr.
 DR Pfam; PF05577; Peptidase_S28; 1.
 DR Hydrolase; Serine protease; Signal.
 FT SIGNAL 1 22 POTENTIAL
 FT CHAIN 23 509 THYMUS-SPECIFIC SERINE PROTEASE.
 FT ACT_SITE 184 184 CHARGE RELAY SYSTEM (POTENTIAL).
 DR

FT ACT SITE 446 446 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 471 471 CHARGE RELAY SYSTEM (POTENTIAL).
 FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 509 AA; 54523 MW; CD8425F3F71EE43C CRC64;
 Query Match 15.3%; Score 73.5; DB 1; Length 509;
 Best Local Similarity 29.2%; Pred. No. 3.4;
 Matches 31; Conservative 13; Mismatches 37; Indels 25; Gaps 5;
 QY 7 VGAGVASADPVADVINTTCYQGVV-ALNATDPG-----AAAQ-----FNASPV 50
 Db 201 VFAAVASSAPLSAVVDFSA-YNQVARSLTQVAIGGSLECLAAASTAFTVERLLRAGA 259
 QY 51 AQSVLRNFLAAPPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
 Db 260 AQAVIREELGA-----CGSLDLEDAELGALQVIGTVQY 297
 RESULT 3
 AMYG_NEUCR STANDARD; PRT; 626 AA.
 ID AMYG_NEUCR STANDARD; PRT; 626 AA.
 AC P14804; Q9P5U5;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Glucoamylase precursor (EC 3.2.1.3) (Glucon 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucohydrolase).
 GN GLA-1 OR B5022.70.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RX MEDLINE=94037144; PubMed=8221928;
 RA Stone P.J., Makoff A.J., Parish J.H., Radford A.;
 RA "Cloning and sequence analysis of the glucoamylase gene of Neurospora
 RA crassa."
 RA Curr. Genet. 24:205-211(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RX MEDLINE=22542210; PubMed=12655011;
 RA Mannhaupt G., Montrone C., Haase D., Mewes H.-W., Aign V.,
 RA Hoheisel J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,
 RA Schulte U.;
 RT "What's in the genome of a filamentous fungus? Analysis of the
 RT Neurospora genome sequence."
 RL Nucleic Acids Res. 31:1944-1954(2003).
 RN [3]
 RP SEQUENCE OF 36-65.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RA Koh-Luar S.I., Parish J.H., Bleasby A.J., Pappin D.J.C., Ainley K.,
 RA Johansen P.E., Radford A.;
 RT "Exported proteins of Neurospora crassa: 1-glucoamylase."
 RL Enzyme Microb. Technol. 11:692-695(1989).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC Glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -!- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X67291; CAA47707.1;
 DR

```

DR EMBL; AL355932; CAB91426.1; -.
DR HSSP; P04064; IGAI.
DR InterPro; IPR002044; CBD 4.
DR InterPro; IPR008291; Glu-a-glcSD_SBD.
DR InterPro; IPR00165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00586; CBM_20; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PIRSF; PIRSF01031; Glu-a-glcSD_SBD; 1.
DR PRINTS; PRO0736; GLHYDRLASE15.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
DR Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal.
FT SIGNAL. 1 19 POTENTIAL.
FT PROPEP 20 35 POTENTIAL.
FT CHAIN 36 626 GLUCOAMYLASE.
FT BINDING 155 155 SUBSTRATE (BY SIMILARITY).
FT ACT_SITE 211 211 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 214 214 GENERAL ACID CATALYST (BY SIMILARITY).
FT ACT_SITE 215 215 INTERACT WITH SUBSTRATE (BY SIMILARITY).
FT CARBOHYD 106 106 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 82 82 MISSING (IN REF. 1).
FT CONFLICT 550 550 A -> R (IN REF. 1).
FT CONFLICT 560 560 V -> L (IN REF. 1).
FT SEQUENCE 626 AA; 66474 MW; 5485BDBBA7A3E349 CRC64;
Query Match 15.2%; Score 73; DB 1; Length 626;
Best Local Similarity 26.9%; Pred. No. 4.8;
Matches 29; Conservative 19; Mismatches 34; Indels 26; Gaps 6;
QY 4 SLTVGAGVASAPVDVAVINTTCNYG----QVVAALNATDGAAGP---NASPVAGSYL- 55
DB 392 SVSTGYSSSTSYTAIINAVTYADGVFDVIAQYTPDGSGLABQFDKDGAPLSATHLT 451
QY 56 ---RNFLAA-----PPQRAAQAQVAGPAAQYIGLVESVAGS 92
DB 452 WSVASFLSAARRAGIVPSWGAASA---NSLPGSCS---ASTVAGS 492
RESULT 4
YD33_MYCTU STANDARD; PRT; 344 AA.
AC Q10674;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Hypothetical protein Rv1333/MT1375/Mb1368.
GN Rv1333 OR MT1375 OR MTCY130.18 OR Mb1368.
OS Mycobacterium tuberculosis, and
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. II, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Akin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -|- SIMILARITY: STRONG TO M.LEPRAE ML1167.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z73902; CAA98097.1; -.
EMBL; AE007011; AAK45639.1; -.
EMBL; BX248338; CAD94229.1; -.
DR PIR; A70771; A70771.
DR MEROPS; T04.001; -.
DR TIGR; MT1375; -.
DR Tuberculist; Rv1333; -.
DR InterPro; IPR005321; Peptidase_S58.
DR Pfam; PF03576; Peptidase_S58; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 161 181- POTENTIAL.
FT TRANSMEM 224 244 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
SQ SEQUENCE 344 AA; 33952 MW; 296A055FECF8E650 CRC64;
Query Match 15.0%; Score 72; DB 1; Length 344;
Best Local Similarity 34.1%; Pred. No. 3.2;
Matches 29; Conservative 7; Mismatches 29; Indels 20; Gaps 3;
QY 2 AMSLTVGAGVASAD-----PVDVAVINTTCNYGVVAALNATDGAAGPAAQNASPVAGSY 54
DB 155 AGALKGVGVTASATLQSGVTGVLAVN-----AAGNVVDPATGLPWAIDLVGEGA 205
QY 55 LRNFLLAAPPPQRAAQAQVAGPAA 79
DB 206 LR---APPAEQIAALALQLSSPLGA 226
RESULT 5
VCAA_BPT7
ID VCAA_BPT7 STANDARD; PRT; 345 AA.
AC P19726; P03177;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Major capsid protein 10A.
GN 10.

```


DR PRINTS; PRO0368; FADPDR.
DR PRINTS; PRO0945; HGRDTASE.
DR PRINTS; PRO0411; PNDRTASEI.
DR PRODOM; PD000139; FAD_Pyr_redox; 1.
DR PROSITE; PS01047; HMA 1; 1.
DR PROSITE; PS0846; HMA 2; 1.
DR PROSITE; PS00076; PYRIDINE REDOX 1; 1.
KW Mercuric resistance; Oxidoreductase; Flavoprotein; FAD; NADP;
KW Mercury; Redox-active center; Metal-binding; Plasmid.
FT DOMAIN 1 66 HMA.
FT NP_BIND 128 136 FAD (ADP PART) (BY SIMILARITY).
FT DISULFID 136 141 REDOX-ACTIVE.
FT NON_CONS 342 343
FT METAL 457 457 MERCURY (POTENTIAL).
FT METAL 458 458 MERCURY (POTENTIAL).
SQ SEQUENCE 460 AA; 48587 MW; 6284666EF3F653F05 CRC64;

Query Match 14.3%; Score 69; DB 1; Length 460;
Best Local Similarity 32.6%; Pred. No. 9.1;
Matches 29; Conservative 14; Mismatches 30; Indels 16; Gaps 5;

QY 9 AGVASADPVDAVNTTCNYGVVAALNATDPGAQAQFNASPVQAQSYLRNFLAAPPQRAA 68
Db 64 AMLADAPPTD---NRTGLEFDKRVGMWGAADKSGAE---RPLQ-----VAVIGSGGAA 110

QY 69 MAAQLQAVPQAQYIGLVE--SVAGSCNN 95
Db 111 MAAALKAVEGQAQ-VTLIERGIGTGVN 138

RESULT 8
REGA RHOSU
ID REGA RHOSU STANDARD; PRT; 183 AA.
AC O82868;
DT 30-MAY-2000 (Rel. 39, Created)
DT 20-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosynthetic apparatus regulatory protein rega.
GN REGA.
OS Rhodovulum sulfidophilum (Rhodobacter sulfidophilus).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodovulum.
OX NCBI_TaxID=35806;
RN [1]
RP SEQUENCE FROM N.A.
RA Masuda S.;
RT "Rhodovulum sulfidophilum photosynthetic regulatory Genes.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM REGB/REGA.
CC INVOLVED IN TRANSACTIVATING ANAEROBIC EXPRESSION OF THE
CC PHOTOSYNTHETIC APPARATUS. IT IS A TRANSCRIPTIONAL REGULATOR THAT
CC IS RESPONSIBLE FOR ACTIVATING EXPRESSION OF THE PUF, PUFH, AND PUC
CC OPERONS IN RESPONSE TO A DECREASE IN OXYGEN TENSION (BY
CC SIMILARITY).
CC -!- PTM: Phosphorylated by regB (Probable).
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AR010722; BAA31474.1; -
CC HSSP; P10958; 1DBW.
DR InterPro; IPR002197; HTH Fis.
DR InterPro; IPR008329; Res_reg_Rega.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF00072; Response_reg; 1.
DR PIRSF; PIRSF036385; Res_reg_Rega; 1.
DR PRODOM; PD000039; Response_reg; 1.

DR SMART: SM00448; REC; 1.
DR TIGRFAMS; TIGR01199; HTH fis; 1.
KW PROSITE; PS0110; RESPONSE REGULATORY; 1.
KW Sensory transduction; Phosphorylation; Transcription regulation;
KW DNA-binding; Activator.
FT DOMAIN 13 127 RESPONSE REGULATORY.
FT MOD_RES 62 62 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 183 AA; 20219 MW; 394DB79D39AAB23F CRC64;

Query Match 14.1%; Score 68; DB 1; Length 183;
Best Local Similarity 27.7%; Pred. No. 4.3;
Matches 28; Conservative 9; Mismatches 32; Indels 32; Gaps 3;

QY 3 MSLTVGAG--VASADP-----VDAVNTTCNYGVVAAL 34
Db 40 MAETVAAGKAIASRPPAYAVVDLRLDGTGLDVVETLRKRPDAKIVLTGYGAIATAV 99

QY 35 NATDPA-----AAQFNASPVQAQSYLRNFLAAPPQRAAMAA 71
Db 100 AAVKVGATDYLKSPADANDVTAALLSNGEALPPPPENPMSA 140

RESULT 9
SRIA PHYPO
ID SRIA PHYPO STANDARD; PRT; 246 AA.
AC P09350;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Spherulin 1A precursor.
OS Eukaryotic polycephalum (Slime mold).
OC Eukaryota; Mycetozoa; Myxogastria; Myxogastromycetidae; Physariida;
OC Physarum.
OX NCBI_TaxID=5791;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88137962; PubMed=2830170;
RA Bernier F.; Lemieux G.; Pallotta D.;
RT "Gene families encode the major encystment-specific proteins of
RT Physarum polycephalum plasmodia.";
RL Gene 59:265-277(1987).
RN [2]
RP PARTIAL SEQUENCE OF 20-34.
RX MEDLINE=93049354; PubMed=1425703;
RA Lane B.G.; Cuming A.C.; Fregeau J.; Carpita N.C.; Hurkman W.J.;
RA Bernier F.; Dratewka-Kos E.; Kennedy T.D.;
RT "Germin isoforms are discrete temporal markers of wheat development.
RT pseudogermin is a uniquely thermostable water-soluble oligomeric
RT protein in ungerminated embryos and like germin in germinated
RT embryos, it is incorporated into cell walls.";
RL Eur. J. Biochem. 209:961-969(1992).
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- DEVELOPMENTAL STAGE: Accumulates specifically during spherulation.
CC -!- MISCELLANEOUS: Spherulin is a major encystment-specific protein.
CC -!- SIMILARITY: Belongs to the germin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M18428; AAA29982.1; -
CC InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin sup.
DR InterPro; IPR001929; Germin.
DR Pfam; PF00190; Cupin; 1.
DR PRINTS; PR00325; GERMIN.
DR PROSITE; PS00725; GERMIN; 1.
KW Cell wall; Signal; Glycoprotein; Manganese; Metal-binding;
KW Multigene family.


```

DR EMBL; U59831; AAC50660.1; -.
DR EMBL; U13222; AAA92039.1; -.
DR PIR; G02738; G02738.
DR PIR; S51627; S51627.
DR HSP; O63245; 2HPH.
DR TRANSFAC; T02472; -.
DR Genew; HGNC:3802; FOXD1.
DR MIM; 601091; -.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR InterPro; IP0001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PRO0053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00339; FORK_HEAD_3; 1.
DR Transcription regulation; DNA-binding; Nuclear protein.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 26 34 POLY-GLU.
FT DOMAIN 39 43 POLY-GLY.
FT DOMAIN 52 57 POLY-ARG.
FT DOMAIN 69 72 POLY-GLU.
FT DOMAIN 73 76 POLY-ASP.
FT DOMAIN 97 113 POLY-GLY.
FT DNA_BIND 124 215 FORK-HEAD.
FT DOMAIN 231 234 POLY-ALA.
FT DOMAIN 252 256 POLY-ALA.
FT DOMAIN 259 266 POLY-PRO.
FT DOMAIN 293 303 POLY-ALA.
FT DOMAIN 309 315 POLY-PRO.
FT DOMAIN 375 390 POLY-ALA.
FT DOMAIN 428 434 POLY-ALA.
SQ SEQUENCE 465 AA; 46140 MW; D3E7854909CCBF6E CRC64;

Query Match 13.9%; Score 67; DB 1; Length 465;
Best Local Similarity 29.0%; Pred. No. 15;
Matches 27; Conservative 15; Mismatches 37; Indels 14; Gaps 4;

QY 8 GAGVASADP--VDVINTCNVGVVVAALNTPGAAQAQFNASPVQAQVFLNFLAAPPPQ 65
DB 353 GASALARSFSTESITGG--SLGPAAAAAQAQAQAQAQASP-----SPVAAPPAP 403

QY 66 RAA--MAAQLOAVPQAQYIGLVESVAGSCNN 95
DB 404 GSSGGCAQAAGVGPAAALTRSLVAAAAAASS 436

RESULT 12
SCWA_YEAST
ID SCWA_YEAST STANDARD; PRT; 389 AA.
AC Q04951;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable family 17 Glucosidase SCW10 precursor (EC 3.2.1.-) (Soluble
DE cell wall protein 10).
GN SCW10 OR YMR305C OR YM9952.07C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagals K., Lye G., Moule S., Odeh C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93(1997).
RN [2]

```

```

RP SEQUENCE OF 30-39, AND SUBCELLULAR LOCATION.
RC STRAIN=SEY6210;
RX MEDLINE=98422453; PubMed=9748433;
RA Cappellaro C., Mrsa V., Tanner W.;
RT "New potential cell wall glucanases of Saccharomyces cerevisiae and
RT their involvement in mating."
RL J. Bacteriol. 180:5030-5037(1998).
CC -!- FUNCTION: Glucanases possibly play a role in cell expansion
CC during growth, in cell-cell fusion during mating, and in spore
CC release during sporulation (by similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- PTM: Glycosylated.
CC -!- SIMILARITY: Belongs to family 17 of glycosyl hydrolases.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; Z49212; CAA89138.1; -.
DR PIR; S53975; S53975.
DR GerMOnline; 142983; -.
DR SGD; S0004921; SCW10.
DR GO; GO:0009277; C:cell wall (sensu Fungi); IDA.
DR GO; GO:0005783; C:endoplasmic reticulum; IDA.
DR GO; GO:0000747; P:conjugation with cellular fusion; IGI.
DR InterPro; IPR000490; Glyco_hydro_17.
DR Pfam; PF00332; Glyco_hydro_17; 1.
DR PROSITE; PS00587; GLYCOSYL_HYDROL_F17; FALSE NEG.
KW Hydrolase; Glycosidase; Cell wall; Glycoprotein; Signal.
FT SIGNAL 1 18
FT PROPEP 19 29
FT CHAIN 30 389 PROBABLE FAMILY 17 GLUCOSIDASE SCW10.
FT DOMAIN 65 134 SER-RICH.
FT ACT_SITE 326 326 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 380 380 PROTON DONOR (BY SIMILARITY).
FT CATHED 279 279 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 38 39 AQ -> QE (IN REF. 2).
SQ SEQUENCE 389 AA; 40469 MW; 21F394BD41137DE1 CRC64;

Query Match 13.8%; Score 66.5; DB 1; Length 389;
Best Local Similarity 27.4%; Pred. No. 14;
Matches 31; Conservative 12; Mismatches 49; Indels 21; Gaps 6;

QY 2 AMSLTVCAG----VASADPVDVINT--TCNYGVVVAALNA---TDPGAAQFNASPVQAQ 52
DB 249 AKSALTSGAGTGVSVSVSDTFIAVINNPDLNCSYSDYM-ANNAHAYFDENTAAQ-DAGPWVL 306

QY 53 SYLRNFLAAPPPQRAAQAOL-----QAQVGAQYIGLVESVAGSCNN 95
DB 307 EQIERVYTAGGKDKDVITETGWPSKGDYGEAVPSKANQEAIAISSIKSSCGS 359

RESULT 13
TSSP_HUMAN
ID TSSP_HUMAN STANDARD; PRT; 514 AA.
AC Q9NQ07; Q7541.6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thymus-specific serine protease precursor (EC 3.4.-.-).
GN PRSS16 OR TSSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RX MEDLINE=99458809; PubMed=10527559;

```

```

RA Bowlus C.L., Ahn J., Chu T., Gruen J.R.;
RT "Cloning of a novel MHC-encoded serine peptidase highly expressed by
RT cortical epithelial cells of the thymus.";
RL Cell. Immunol. 196;80-86(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Phillips S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Protease that may play a role in T-cell development.
CC -!- SUBCELLULAR LOCATION: Vesicular, either lysosomal or endosomal.
CC -!- TISSUE SPECIFICITY: Expressed predominantly in cortical thymic
CC epithelial cells.
CC -!- DEVELOPMENTAL STAGE: Expressed in fetal thymus.
CC -!- SIMILARITY: Belongs to peptidase family S28.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF052514; AAC33563.1; -
CC EMBL; AL021808; CAB94769.1; ALT_SEQ.
CC MEROPS; S28.003; -.
CC DR Genew; HGNC:9480; PRSS16.
CC DR MIM; 607169; -.
CC DR GO; GO:0008236; F:serine-type peptidase activity; NAS.
CC DR GO; GO:0030163; P:protein catabolism; NAS.
CC DR InterPro; IPR008758; Peptidase_S28.
CC DR Pfam; PF05577; Peptidase_S28; 1.
CC KW Hydrolase; Serine protease; Signal.
CC FT SIGNAL 1 24
CC FT CHAIN 25 514 THYMUS-SPECIFIC SERINE PROTEASE.
CC FT ACT_SITE 185 185 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 447 447 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 472 472 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT CARBOHYD 70 70 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 172 172 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 321 321 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 514 AA; 55048 MW; EF39BC06C560709A CRC64;

Query Match 13.8%; Score 66.5; DB 1; Length 514;
Best Local Similarity 31.8%; Pred.No.19;
Matches 31; Conservative 14; Mismatches 30; Indels 23; Gaps 5;

QY 9 AGVASADPVDVAVINTTCYGVVVAALNA-ALNATDFGAAQAQFNAS-----PVAQ 52
Db 204 ASVASSAPRAVLDFE-EYNDVVSRSLMSTALGSLDCEAAVSVAPAEVERLRSGGAAQ 262

QY 53 SYLENFLAPPP-----QRAAMAQLOA-VFGNAQYIG 84
Db 263 AALRTLSACGPIGRAENQAEALLGALQALVGGVVQYDQ 300

RESULT 14
DCOS_ECOLI STANDARD; PRT; 732 AA.
AC P24169;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ornithine decarboxylase, inducible (EC 4.1.1.17).
GN SPEG OR B0693.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92041955; PubMed=1939141;
RA Kashiwagi K., Suzuki T., Furuchi T., Kobayashi H.,

```

```

RA Igarashi K.;
RT "Coexistence of the genes for putrescine transport protein and
RT ornithine decarboxylase at 16 min on Escherichia coli chromosome.";
RL J. Biol. Chem. 266:20922-20927(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayaishi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -!- CATALYTIC ACTIVITY: L-ornithine = putrescine + CO(2).
CC -!- COFACTOR: Pyridoxal phosphate.
CC -!- PATHWAY: Biosynthesis of spermidine from arginine; third step.
CC -!- INDUCTION: By low environmental pH.
CC -!- SIMILARITY: Belongs to the Orn/Lys/Arg decarboxylase class-I
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M64495; AAA62785.1; -.
CC EMBL; AF000172; AAC73787.1; -.
CC EMBL; D90708; BAA35349.1; -.
CC PIR; A40839; A40839.
CC HSSP; P43099; 1C4K.
CC EcGene; EG10964; speF.
CC InterPro; IPR000310; Decarboxylase.
CC InterPro; IPR008286; Decarboxylase_C.
CC InterPro; IPR005308; OKR_DC_1_N.
CC Pfam; PF01276; OKR_DC_1_1.
CC Pfam; PF03711; OKR_DC_1_C; 1.
CC Pfam; PF03709; OKR_DC_1_N; 1.
CC PROSITE; PS00703; OKR_DC_1; 1.
CC Spermidine biosynthesis; Lyase; Decarboxylase; Pyridoxal phosphate;
KW Complete proteome.
SQ BINDING 355 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 732 AA; 82415 MW; C2C030BEAB9D23F6 CRC64;

Query Match 13.8%; Score 66.5; DB 1; Length 732;
Best Local Similarity 26.0%; Pred.No.27;
Matches 34; Conservative 12; Mismatches 38; Indels 47; Gaps 5;

QY 9 AGVASADPVDVAVINTTCYGVVVAALNA-----TDPGAAQAQFNASPV- 50
Db 184 AKVENADKTYFVLNGTSSNNKV--LNALLTPGDLVLFDRNHNKSHHGALLQAGATPVY 241

QY 51 -----AQSYLRFNLAAPPORAMAAQLO-AVPGAQYIG----- 84
Db 242 LETARNPYFGIGDAHCFEESYLRELIAEVAPQRAKEAPFLAVIQLGTGYDGTYNAR 301

QY 85 -LVESVAGSCN 94

```

Db 302 QVVDKIGHLCD 312

RESULT 15

VGLH_HSVBC

ID_VGLH_HSVBC STANDARD; PRT; 842 AA.

AC P27539;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 46, Last annotation update)

DE Glycoprotein H precursor.

GN GH OR UL22.

OS Bovine herpesvirus type 1 (strain Cooper).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicellovirus.

OX NCBI_TaxID=10323;

RN [1]

RP SEQUENCE FROM N.A.

RA Meyer A.L., Petrovskis E.A., Duffus W.P.N., Thomsen D.R.,

RA Post L.E.;

RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: GH AND GL FORM A COMPLEX THAT INDUCES NEUTRALIZING AND

CC SYNCYTIA INHIBITING ANTIBODIES (BY SIMILARITY).

CC -!- SIMILARITY: Belongs to the herpesviruses glycoprotein H family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X58867; CAA41677.1; -.

DR EMBL; J78205; CAB01604.1; -.

DR EMBL; AJ004801; CAA06112.1; -.

DR PIR; S18462; S18462.

KW Glycoprotein; Transmembrane; Signal.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 842 GLYCOPROTEIN H.

FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 617 617 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 783 783 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 842 AA; 88375 MW; BA90759A74715F98 CRC64;

Query Match 13.8%; Score 66.5; DB 1; Length 842;

Best Local Similarity 28.4%; Pred. No. 31;

Matches 25; Conservative 11; Mismatches 43; Indels 9; Gaps 3;

QY 15 DPVDAVINTCNYG--QVVAALNATDGAAGFNA---SPVAQSYLRNFLAAPPORA 67

Db 447 EPAGAALESVVARGKRLRAVAFSGDPAPLAARLLAHGVVSDLYDAFLRGELTWGPFNRH 506

QY 68 AM--AAQLQAVPAAQYIGLVESVAGSC 93

Db 507 ALFFFAVAASAFPADAQALELARDVTRKC 534

Search completed: June 22, 2004, 17:17:00

Job time : 3.68859 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 7.84949 Seconds
(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349A-22
Perfect score: 481
Sequence: 1 VAMSLTVGAGVASADPVDV.....PCAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriapi.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	481	100.0	110	050430	mycobacteri
2	481	100.0	110	0700G8	mycobacteri
3	79	16.4	1243	09W6T6	Q9w6t6 brachydanio
4	78	16.2	280	08S2D6	Q8s2d6 oryza sativ
5	77	16.0	410	08RZ79	Q8rz79 oryza sativ
6	75.5	15.7	336	098HE8	Q98he8 rhizobium 1
7	75	15.6	337	08NTY7	Q8nty7 corynebacte
8	74	15.4	586	08LN25	Q8ln25 oryza sativ
9	74	15.4	2213	08XN25	Q8xn25 chlamydomon
10	73.5	15.3	462	08VPA4	Q8vpa4 pseudomonas
11	73.5	15.3	466	08DWE7	Q8dwe7 streptococc
12	73.5	15.3	1489	08DWP5	Q8dwp5 agrobacteri
13	72.5	15.1	190	084JP1	Q84jp1 arabidopsis
14	72.5	15.1	197	09S9P9	Q9s9p9 arabidopsis
15	72.5	15.1	654	08P38	Q8p38 xanthomonas
16	72	15.0	143	083VK1	Q83vk1 sphingomonas

17	72	15.0	196	16	09JW92	Q9jw92 neisseria m
18	72	15.0	241	16	09K313	Q9k313 streptomyce
19	71.5	14.9	308	12	039491	Q39491 bovine herp
20	71	14.8	314	5	08INP3	Q8inp3 drosophila
21	71	14.8	372	5	08INP4	Q8inp4 drosophila
22	71	14.8	441	16	08XQP8	Q8xqp8 raietonia s
23	71	14.8	516	16	07UPE9	Q7upe9 rhodopirell
24	71	14.8	849	12	09WQY5	Q9wqy5 bovine herp
25	70.5	14.7	222	2	093L09	Q93l09 burkholderi
26	70.5	14.7	249	5	09UIN3	Q9uin3 cryptospori
27	70.5	14.7	462	16	09HWK6	Q9hwk6 pseudomonas
28	70.5	14.7	555	5	090625	090625 cryptospori
29	70.5	14.7	618	11	091221	Q91221 mus musculu
30	70.5	14.7	689	11	09DBI6	Q9dbi6 mus musculu
31	70.5	14.7	689	11	08CZQ3	Q8czq3 mus musculu
32	70.5	14.7	1832	5	096503	Q96503 cryptospori
33	70.5	14.7	1850	5	09GRJ7	Q9grj7 leishmania
34	70	14.6	179	5	09VW18	Q9vw18 drosophila
35	70	14.6	196	16	09JXN1	Q9jxn1 neisseria m
36	70	14.6	293	5	017222	017222 caenorhabdi
37	70	14.6	411	16	08G431	Q8g431 bifidobacte
38	70	14.6	414	16	08PPX6	Q8ppx6 xanthomonas
39	70	14.6	876	4	08WY43	Q8wy43 homo sapien
40	70	14.6	1064	4	09HAN2	Q9han2 homo sapien
41	70	14.6	7525	2	09KIE0	Q9kie0 streptomyce
42	69.5	14.4	273	16	07UKL8	Q7ukl8 rhodopirell
43	69.5	14.4	288	17	09HNP1	Q9hnp1 halobacteri
44	69.5	14.4	728	16	08XQ78	Q8xq78 raietonia s
45	69.5	14.4	759	10	094KS5	Q94ks5 hordeum vul

ALIGNMENTS

RESULT 1

050430	ID	050430	PRELIMINARY;	PRT;	110 AA.
AC	050430;				
DT	01-JUN-1998	(TREMBlrel. 06; Created)			
DT	01-JUN-1998	(TREMBlrel. 06; Last sequence update)			
DT	01-JUN-2003	(TREMBlrel. 24; Last annotation update)			
DE	Hypothetical protein Rv1174c.				
OS	Rv1174C OR MTV005.10C OR MT1211.				
OC	Mycobacterium tuberculosis.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,				
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;				
RA	Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,				
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,				
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,				
RA	Bishai W.;				
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and				
RT	laboratory strains."				
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				

```

DR EMBL; AL010186; CAA15851.1; -.
DR EMBL; AE006998; AAK45468.1; -.
DR PIR; F70875; F70875.
DR TIGR; MT1211; -.
DR Tuberculist; Bv1174c; -.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 110 AA; 10881 MW; 0309DDB489150061 CRC64;

Query Match 100.0%; Score 481; DB 16; Length 110;
Best Local Similarity 100.0%; Pred. No. 3.1e-39;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGGAAQFNASPVASQYLNFIA 60
Dy 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGGAAQFNASPVASQYLNFIA 74
Qy 61 APPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 96
Dy 75 APPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 110

RESULT 2
Q7U0G8 PRELIMINARY; PRT; 110 AA.
AC Q7U0G8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Hypothetical protein.
GN MB1207C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248337; CAD94068.1; -.
KW Complete proteome.
SQ SEQUENCE 110 AA; 10881 MW; 0309DDB489150061 CRC64;

Query Match 100.0%; Score 481; DB 16; Length 110;
Best Local Similarity 100.0%; Pred. No. 3.1e-39;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGGAAQFNASPVASQYLNFIA 60
Dy 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGGAAQFNASPVASQYLNFIA 74
Qy 61 APPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 96
Dy 75 APPQRAAMAALQAVPGAAQYIGLVESVAGSCNNY 110

RESULT 3
Q9W6T6 PRELIMINARY; PRT; 1243 AA.
AC Q9W6T6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Patched-2 protein.
GN PTC2 OR PTC1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;

```

```

OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Lewis K.E., Concordet J.P., Ingham P.W.;
RT "Characterisation of the second ptc gene in zebrafish.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ007742; CAB39726.1; -.
DR ZFIN; ZDB-GENE-980526-196; ptc2.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008158; P:hedghog receptor activity; IEA.
DR InterPro; IPR003392; Patched.
DR InterPro; IPR004766; Patched_rec.
DR InterPro; IPR000731; SSD_5TM.
DR Pfam; PF02460; Patched; 1.
DR TIGRFAMs; TIGR00918; 2A060602; 1.
DR PROSITE; PS50156; SSD; 1.
SQ SEQUENCE 1243 AA; 138548 MW; FABF459DDBOC91371 CRC64;

Query Match 16.4%; Score 79; DB 13; Length 1243;
Best Local Similarity 29.3%; Pred. No. 49;
Matches 27; Conservative 11; Mismatches 40; Indels 14; Gaps 3;

Qy 18 DAVINTTCNYGVVAALNATDP--GAAQFNASPVASQYLNR-----NFLAAPP 64
Dy 925 DGIINPNAFY-IYLSAVNSDPVAYAAQNIIRHPPEWLHRTDSIPASRLNIPAAEPI 983
Qy 65 QRAAMAAALQAVPGAAQYIGLVESVAGSCNNY 96
Dy 984 EYAQPPFYINGLRETPOFVEAIESVRAICNNY 1015

RESULT 4
Q8S2D6 PRELIMINARY; PRT; 280 AA.
AC Q8S2D6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P0401G10.4 protein (Similar to RNA-binding protein).
GN P0401G10.4 OR Q1294_F06.21.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
clone:P0401G10.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone:QJ1294_F05.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003238; BAB88997.1; -.
DR EMBL; AP004326; BAB92880.1; -.
DR Gramene; Q8S2D6; -.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; xrm; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 280 AA; 28689 MW; 852687706D3DBC5F CRC64;

Query Match 16.2%; Score 78; DB 10; Length 280;
Best Local Similarity 34.2%; Pred. No. 12;
Matches 27; Conservative 6; Mismatches 42; Indels 4; Gaps 2;

```

QY 9 AGVASADPV-DVAVINTTCNGVVAALNATDPGAAQFNASPVASQYLRFNLAAPPORA 67
 DB 198 AGLTWANTGGGLSPYFYQYGPVAVANAAAGYAGMQY---PQMYQYVAAAAAAGCAATA 244
 QY 68 AMAAQLOAVPQAAQYIGLV 86
 DB 245 TTASQLTAVAGLQVAGTV 263

RESULT 5
 Q8R279 PRELIMINARY; PRT; 410 AA.
 ID Q8R279
 AC Q8R279
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE B1065G12.16 protein.
 GN B1065G12.16
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
 clone: B1065G12.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBSJ databases.
 CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 DR EMBL; AP003791; BAB90534.1; -.
 DR Gramene; Q8R279; -.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR002018; CarbohydraseB.
 DR InterPro; IPR000379; Ser_estr.
 DR Pfam; PF00135; Coesterase; 1.
 KW Hydrolase.
 SQ SEQUENCE 410 AA; 44077 MW; 64F2F2C2B5E1D3B7 CRC64;

Query Match 16.0%; Score 77; DB 10; Length 410;
 Best Local Similarity 28.3%; Pred. No. 22;
 Matches 28; Conservative 12; Mismatches 49; Indels 10; Gaps 3;

QY 4 SLTVGAGVASADPVD--AVINTTCNGVQVVA---ALNATDPGAAQFNASPVASQYLRFN 57
 DB 23 SLAAAAAAAAGGEDDAAGFVGSRADEATAPSNPAFSAADGVASKDLHIDENSSLSVRI 82
 QY 58 FLAAPP-----QPAAMAAQLQAVPGAAQYIGLVESVAGS 92
 DB 83 FLUPTPPPHSRRASEPPPPATANGSGAPYRGYLDPHAVSS 121

RESULT 6
 Q8R279 PRELIMINARY; PRT; 336 AA.
 ID Q8R279
 AC Q8R279
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Flagellar motor switch protein, Flig.
 GN M12901.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214958;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Idesawa K., Ishikawa K., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RT DNA Res. 7:331-338(2000).
 RL EMBL; AP003000; BAB49918.1; -.
 DR GO; GO:000288; C:flagellum (sensu Bacteria); IEA.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0006335; F:structural constituent of ribosome; IEA.
 DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR000090; Flg_Motor_Flig.
 DR InterPro; IPR000307; Ribosomal_S16.
 DR Pfam; PF01706; Flig-C; 1.
 DR PRINTS; PR00954; FLGMOTORFLIG.
 DR PROSITE; PS00732; RIBOSOMAL_S16; 1.
 KW Complete proteome.
 SQ SEQUENCE 336 AA; 35986 MW; 90FCFA5ED7B0FEDB CRC64;

Query Match 15.7%; Score 75.5; DB 16; Length 336;
 Best Local Similarity 30.6%; Pred. No. 25;
 Matches 26; Conservative 13; Mismatches 39; Indels 7; Gaps 2;

QY 6 TVGAGVA-SADPVDVAVINTTCNGVQVVAALNATDPGAAQFNASPVASQYLRFN 58
 DB 67 TEGALLSDADMDTILNLSPEMSAIMGNKKPEAAPEGPPPIWDLKLEPSRLGTF 126
 QY 59 LAAPPPORAAMAAQLQAVPGAAQYI 83
 DB 127 LAGEHPOTAAVLSKLAPQAAASVL 151

RESULT 7
 Q8NTY7 PRELIMINARY; PRT; 337 AA.
 ID Q8NTY7
 AC Q8NTY7
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Predicted dehydrogenases and related proteins (EC 1.1.1.18).
 GN CGL0164.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AP005274; BAB97557.1; -.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000683; GFO_IDH_MocA.
 DR InterPro; IPR004104; GFO_IDH_MocA_C.
 DR Pfam; PF01408; GFO_IDH_MocA; 1.
 DR Pfam; PF02894; GFO_IDH_MocA_C; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 337 AA; 36431 MW; F5EF6EAF6A5924A3 CRC64;

Query Match 15.6%; Score 75; DB 16; Length 337;
 Best Local Similarity 37.3%; Pred. No. 28;
 Matches 28; Conservative 11; Mismatches 26; Indels 10; Gaps 4;

QY 7 VGAGVASADPVDVAVINTTCNGVQVVAALNATDPGAAQFNASPVASQYLRFNLAAPP 63
 DB 10 VGAGAGADHIDRINNRTS--GAHISAIIEPDAFAAAAAEDAPGAQAFTRIEDAIAAD- 66

65 AAPARRVARAAAPLAPKPGTBLQGV 90

DB

RESULT 11

QSDWE7

ID QSDWE7 PRELIMINARY; PRT; 466 AA.

AC QSDWE7

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Putative PTS system, fructose-specific IIBC component.

GN SMU.114.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=UA159 / ATCC 700610 / Serotype C;

RX MEDLINE=2229563; PubMed=12397186;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J., Qian Y., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.; "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).

RL EMBL; AB014863; JN57895.1; --

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005351; F:sugar porter activity; IEA.

DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.

DR InterPro; IPR003352; P:trans_EIIC.

DR InterPro; IPR003353; P:trans_IIB_fruct.

DR InterPro; IPR006327; PTS_IIC_fruct.

DR Pfam; PF02378; PTS_EIIC; 1.

DR Pfam; PF02379; PTS_IIB_fruct; 1.

DR TIGRFAMs; TIGR00829; FRU; 1.

DR TIGRFAMs; TIGR01427; PTS_IIC_fruct; 1.

KW Complete proteome.

SQ SEQUENCE 466 AA; 48902 MW; 3C7244E497A95815 CRC64;

Query Match 15.3%; Score 73.5; DB 16; Length 466;

Best Local Similarity 26.0%; Pred. No. 56;

Matches 25; Conservative 16; Mismatches 46; Indels 9; Gaps 2;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNGYGVAAALNATDFGAAQFNASPVASQYLRFNFA 60

DB 150 VLVAISFLGIYGFDPKSYQYH-----FAALLKTDIGGVAMGWTILAAIYAESIA 202

QY 61 APPPPRAAQAQLOAVPFGAAQYIGLVESVAGSCNNY 96

DB 203 KRPGFVAGFVGGMVANGSGFLGGI--TAGFAAGY 236

RESULT 12

Q8U9P5

ID Q8U9P5 PRELIMINARY; PRT; 1489 AA.

AC Q8U9P5

DT 01-JUN-2002 (T-EMBLrel. 21, Created)

DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Polyketide synthetase.

GN ATU3681 OR AGR_L_2313.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.

OX NCBI_TaxID=176299;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P., Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L., Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr., Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,

Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Raymond C., Rouse C., Saenphimmachak C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-X., Dolan M., Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V., Nester E.W.;

RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58.";

RT Science 294:2317-2323 (2001).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608551; PubMed=11743194;

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Ouello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L., Hummel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F., Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B., Flanagan C., Crowell C., Gurson J., Lono C., Sear C., Strub G., Cielo C., Slater S.;

RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58.";

RT Science 294:2323-2328 (2001).

RL EMBL; AE009297; AAL44493.1; ALT_INIT.

DR EMBL; AE008316; AAK89721.1; --

DR PIR; AG3009; AG3009.

DR PIR; G98274; G98274.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.

DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR001227; AC:trans.

DR InterPro; IPR00794; Ketoacyl synth.

DR InterPro; IPR006162; P:antenn_S.

DR InterPro; IPR006163; P:bind.

DR Pfam; PF00698; Acyl transf; 1.

DR Pfam; PF00109; ketoacyl-synt; 1.

DR Pfam; PF02801; ketoacyl-synt; 1.

DR Pfam; PF00550; PP-binding; 1.

DR PROSITE; PS00606; ACP_DOMAIN; 1.

DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.

KW Complete proteome.

SQ SEQUENCE 1489 AA; 159121 MW; 5BF4FF9FARCC90009 CRC64;

Query Match 15.3%; Score 73.5; DB 16; Length 1489;

Best Local Similarity 28.6%; Pred. No. 2.1e+02;

Matches 26; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

QY 5 LTVGAGVASADPVDVAVINTTCNGYGVAAALNATDFGAAQFNASPVA-OSYLRFNFAA 61

DB 373 LDTAAGIA-----GFIKTVLCKHGRIPASLHFRQANAKIDFPASPFAVNTVLRDWTAL 426

QY 62 PFPORAAQAQLOAVPFGAAQYIGLVESVAGS 92

DB 427 ESPRRAGISS--LGIGGTNAHVLEAPAGA 455

RESULT 13

Q84JPI

ID Q84JPI PRELIMINARY; PRT; 190 AA.

AC Q84JPI

DT 01-JUN-2003 (TREMELrel. 24, Created)

DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Putative transcription factor.

GN AtUG30500.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608550; PubMed=11743193;

RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,

DR EMBL; AC009917; AAF19754.1; -.
DR PIR; A86430; A86430.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001289; TF_CBEF.
DR Pfam; PF02045; CBEF_NFYA; 1.
DR PRINTS; PR00616; CCAATSUBUNTB.
DR ProDom; PD003860; TF_CBEF; 1.
DR SMART; SM00521; CBF; 1.
SQ SEQUENCE 197 AA, 22127 MW; EB67C5050BE0F874 CRC64;

Query Match 15.1%; Score 72.5; DB 10; Length 197;
Best Local Similarity 31.0%; Pred. No. 27;
Matches 26; Conservative 8; Mismatches 23; Indels 27; Gaps 96

QY 26 NYGQVVAAALNATDPG-----AAACFNASPVQASVLYRFLAAPPDQ-----RAKARALQ 74
DB 41 NYESVTSLVYSDPQTTSMA PQGQ---PYDDPYRS-IFAPDPQYTGVLQLMGVQQ 96

QY 75 AVP-----GAAQYIGLV 86
DB 97 GVPLPSDAVEPVPVFNQKQYHGIL 120

RESULT 15
Q8P938 PRELIMINARY; PRT; 654 AA.
ID Q8P938 AC Q8P938;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein XCC2028.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorlo C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.W., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Clapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Forghieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Maidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.;
RRL Nature 417:459-463(2002).
RRL EMBL; AEO12308; AAM41317.1; -.
DR GO; GO:0004563; F:beta-N-acetylhexosaminidase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001540; Glyco_hydro_20.
DR Pfam; PF02838; Glyco_hydro_20b; 1.
DR HypoProteome; Complete proteome.
SQ SEQUENCE 654 AA; 69164 MW; 9A96E86E1302F410 CRC64;

Query Match 15.1%; Score 72.5; DB 16; Length 654;
Best Local Similarity 29.9%; Pred. No. 1e+02;
Matches 20; Conservative 11; Mismatches 21; Indels 15; Gaps 1

Db	588	VSAALANSQATQOFDAA-----KQLAEQAAAIQSIPGATRFAGPVKVA	632
Qy	90	AGSCNY	96
Db	633	DGVLSF	639

Search completed: June 22, 2004, 17:22:57
Job time : 10.8495 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 55 Seconds
(without alignments)
498.311 Million cell updates/sec

Title: US-09-886-349A-24

Perfect score: 468

Sequence: 1 MSLDAHPQLVASQSAFAA.....EAGTYVAADAAAATYGTG 97

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Genesep29Jan04:*

- 1: Genesep1980s:*
- 2: Genesep1990s:*
- 3: Genesep2000s:*
- 4: Genesep2001s:*
- 5: Genesep2002s:*
- 6: Genesep2003as:*
- 7: Genesep2003bs:*
- 8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	100.0	97	2	AAW73653 M. tuberc
2	468	100.0	97	2	AAW73759 M. tuberc
3	468	100.0	97	2	AAW73763 M. tuberc
4	468	100.0	97	2	AAW73869 M. tuberc
5	468	100.0	97	4	AB35221 M. tubercu
6	468	100.0	97	4	AAU08229 Mycobacte
7	468	100.0	97	4	AAU08209 Mycobacte
8	468	100.0	97	4	AAU01893 M. tuberc
9	468	100.0	97	5	AAE29712 Mycobacte
10	468	100.0	97	5	AAE17576 Mycobacte
11	468	100.0	299	5	AAE32067 Mycobacte
12	468	100.0	299	5	AAE29723 Mycobacte
13	468	100.0	299	5	AAE17587 Mycobacte
14	468	100.0	306	5	AAU74596 Antigenic
15	468	100.0	433	5	AAE32065 Mycobacte
16	468	100.0	433	5	AAU74594 Antigenic
17	468	100.0	710	5	AAE32066 Mycobacte
18	468	100.0	710	5	AAE29724 Mycobacte
19	468	100.0	710	5	AAE17588 Mycobacte
20	468	100.0	723	5	AAU74595 Antigenic
21	468	100.0	856	2	AAE32064 Mycobacte
22	468	100.0	859	5	AAU74593 Antigenic
23	442	94.4	97	4	AB35228 M. tubercu
24	352	75.2	98	5	ABU05649 M. tuberc
25	90.5	19.3	104	7	ABE10630 Structura

26	89.5	19.1	405	6	ABU34103 Protein e
27	88.5	18.9	163	6	ABU34018 Protein e
28	87.5	18.7	95	2	AAW32376 Mycobacte
29	87.5	18.7	95	2	AAW32444 Mycobacte
30	87.5	18.7	95	2	AAW64321 Mycobacte
31	87.5	18.7	95	2	AAW81747 M. tuberc
32	87.5	18.7	95	2	AAW32097 Mycobacte
33	87.5	18.7	95	2	AAW38981 M. tuberc
34	87.5	18.7	95	2	AAW39118 M. tuberc
35	87.5	18.7	95	5	AAE29717 Mycobacte
36	87.5	18.7	95	5	AAE17581 Mycobacte
37	87.5	18.7	100	2	AAW64339 Mycobacte
38	87.5	18.7	100	2	AAW81706 M. tuberc
39	87.5	18.7	100	2	AAW32097 M. tuberc
40	87.5	18.7	100	2	AAW38993 M. tuberc
41	87.5	18.7	100	2	AAW39136 M. tuberc
42	87.5	18.7	100	4	AAW19845 Mycobacte
43	87.5	18.7	100	4	AAW35218 M. tubercu
44	87.5	18.7	100	5	ABG30969 Mycobacte
45	87.5	18.7	100	5	ABU05988 M. tuberc

ALIGNMENTS

RESULT 1
AAW73653
ID AAW73653 standard; protein; 97 AA.
XX
AC AAW73653;
XX
DT 24-MAR-1999 (first entry)
XX
DE M. tuberculosis antigen clone Tb472 protein sequence.
XX
KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis; infection.
XX
OS Mycobacterium tuberculosis.
XX
FN W09853076-A2.
XX
PD 26-NOV-1998.
XX
PF 20-MAY-1998; 98WO-US010514.
XX
PR 20-MAY-1997; 97US-00858998.
PR 05-MAY-1998; 98US-00073009.
XX
XX (CORI-) CORIXA CORP.
XX
XX Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX
XX WPI; 1998-045315/04.
XX
XX N-PSDB; AAX01132.
XX
XX New isolated Mycobacterium tuberculosis antigens - used to develop products for the prevention, treatment and diagnosis of tuberculosis infection.
XX
XX Claim 2; Page 76-77; 104pp; English.

This sequence represents an immunogenic portion of a Mycobacterium tuberculosis antigen of the invention. The polypeptides are useful for immunotherapy to treat or prevent tuberculosis (especially in humans), e.g. they can be included with an acceptable carrier in pharmaceutical compositions or included in vaccines, and administered to induce protective immunity in a patient against M. tuberculosis. Tuberculosis is a chronic, infectious disease generally caused by M. tuberculosis infection, and if left untreated typically results in serious complications and death. Fusion proteins containing the antigen, or DNA molecules can similarly be included with an acceptable carrier in pharmaceutical compositions or in vaccines and administered as above. The

CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
 CC dermal cells with at least one polypeptide and detecting an immune
 CC response (especially induration) on the patient's skin. Inhibiting the
 CC spread of tuberculosis requires vaccination and accurate diagnosis, since
 CC antibiotic therapy may not be effective due to the existence of an
 CC asymptomatic but contagious stage and to patient non-compliance. The
 CC polypeptides overcome concerns of safety and efficacy of current
 CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 CC lack of sensitivity and specificity of existing diagnostic techniques
 XX
 XX Sequence 97 AA;

Query Match 100.0%; Score 468; DB 2; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAFQAAWSAQAFHQGESSAAFOAAHARFVA 60
 DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAFQAAWSAQAFHQGESSAAFOAAHARFVA 60
 QY 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
 DB 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

RESULT 2
 AAW73759
 ID AAW73759 standard; peptide; 97 AA.
 XX
 AC AAW73759;
 XX
 DT 24-MAR-1999 (first entry)
 DE
 DE M. tuberculosis antigen clone protein #1.
 XX
 KW Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 KW infection.
 XX Mycobacterium tuberculosis.
 CS
 XX WO9853076-A2.
 FN
 XX 26-NOV-1998.
 PD
 XX 20-MAY-1998; 98WO-US010514.
 PF
 XX 20-MAY-1997; 97US-00858998.
 PR
 XX 05-MAY-1998; 98US-00073009.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
 PI
 XX WPI; 1999-045315/04.
 DR
 XX New isolated Mycobacterium tuberculosis antigens - used to develop
 PT products for the prevention, treatment and diagnosis of tuberculosis
 PT infection.
 XX
 XX Disclosure; Page 92; 104pp; English.

CC This sequence represents an immunogenic portion of a Mycobacterium
 CC tuberculosis antigen of the invention. The polypeptides are useful for
 CC immunotherapy to treat or prevent tuberculosis (especially in humans),
 CC e.g. they can be included with an acceptable carrier in pharmaceutical
 CC compositions or included in vaccines, and administered to induce
 CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
 CC a chronic, infectious disease generally caused by M. tuberculosis
 CC infection, and if left untreated typically results in serious
 CC complications and death. Fusion proteins containing the antigen, or DNA
 CC molecules can similarly be included with an acceptable carrier in
 CC pharmaceutical compositions or in vaccines and administered as above. The
 CC polypeptides are also useful for diagnosis of tuberculosis, by contacting

CC dermal cells with at least one polypeptide and detecting an immune
 CC response (especially induration) on the patient's skin. Inhibiting the
 CC spread of tuberculosis requires vaccination and accurate diagnosis, since
 CC antibiotic therapy may not be effective due to the existence of an
 CC asymptomatic but contagious stage and to patient non-compliance. The
 CC polypeptides overcome concerns of safety and efficacy of current
 CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 CC lack of sensitivity and specificity of existing diagnostic techniques
 XX
 XX Sequence 97 AA;

Query Match 100.0%; Score 468; DB 2; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAFQAAWSAQAFHQGESSAAFOAAHARFVA 60
 DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAFQAAWSAQAFHQGESSAAFOAAHARFVA 60
 QY 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
 DB 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

RESULT 3
 AAW73763
 ID AAW73763 standard; protein; 97 AA.
 XX
 AC AAW73763;
 XX
 DT 24-MAR-1999 (first entry)
 DE
 DE M. tuberculosis antigen clone Tb472 protein sequence.
 XX
 KW Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 KW infection.
 XX Mycobacterium tuberculosis.
 CS
 XX WO9853075-A2.
 FN
 XX 26-NOV-1998.
 PD
 XX 20-MAY-1998; 98WO-US010407.
 PF
 XX 20-MAY-1997; 97US-00859381.
 PR
 XX 05-MAY-1998; 98US-00073010.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
 PI
 XX WPI; 1999-045314/04.
 DR
 XX N-PSDB; AAX01166.
 DR
 XX Polypeptide comprising immunogenic Mycobacterium tuberculosis antigen -
 PT useful for immunisation against M. tuberculosis infection to treat or
 PT prevent tuberculosis, and in diagnosis of tuberculosis.
 XX
 XX Claim 2; Page 75-76; 100pp; English.

CC This sequence represents an immunogenic portion of a Mycobacterium
 CC tuberculosis antigen of the invention. The polypeptides are useful for
 CC immunotherapy to treat or prevent tuberculosis (especially in humans),
 CC e.g. they can be included with an acceptable carrier in pharmaceutical
 CC compositions or included in vaccines, and administered to induce
 CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
 CC a chronic, infectious disease generally caused by M. tuberculosis
 CC infection, and if left untreated typically results in serious
 CC complications and death. Fusion proteins containing the antigen, or DNA
 CC molecules can similarly be included with an acceptable carrier in
 CC pharmaceutical compositions or in vaccines and administered as above. The
 CC polypeptides are also useful for diagnosis of tuberculosis, by contacting

CC dermal cells with at least one polypeptide and detecting an immune
CC response (especially induration) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually *Bacillus Calmette-Guerin*) and
CC lack of sensitivity and specificity of existing diagnostic techniques
XX
SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.3e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMGAQAFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMGAQAFHQGESSAAAFQAAHARFVA 60

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97

RESULT 4
AAW73869
ID AAW73869 standard; peptide; 97 AA.
AC AAW73869;
DT 24-MAR-1999 (first entry)
DE M. tuberculosis antigen cloneb protein #1.
KW Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
KW infection.
OS Mycobacterium tuberculosis.
PN WO9853075-A2.
XX 26-NOV-1998.
XX 20-MAY-1998; 98WO-US010407.
XX 20-MAY-1997; 97US-00859381.
XX 05-MAY-1998; 98US-00073010.
XX (CORI-) CORIXA CORP.
PA
PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX
XX WPI; 1999-045314/04.
XX
XX Polypeptide comprising immunogenic Mycobacterium tuberculosis antigen -
PT useful for immunisation against M. tuberculosis infection to treat or
PT prevent tuberculosis, and in diagnosis of tuberculosis.
XX
PS Disclosure; Page 91; 100pp; English.
XX

This sequence represents an immunogenic portion of a Mycobacterium
CC tuberculosis antigen of the invention. The polypeptides are useful for
CC immunotherapy to treat or prevent tuberculosis (especially in humans),
CC e.g. they can be included with an acceptable carrier in pharmaceutical
CC compositions or included in vaccines, and administered to induce
CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC a chronic, infectious disease generally caused by M. tuberculosis
CC infection, and if left untreated typically results in serious
CC complications and death. Fusion proteins containing the antigen, or DNA
CC molecules can similarly be included with an acceptable carrier in
CC pharmaceutical compositions or in vaccines and administered as above. The
CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC dermal cells with at least one polypeptide and detecting an immune

CC response (especially induration) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually *Bacillus Calmette-Guerin*) and
CC lack of sensitivity and specificity of existing diagnostic techniques
XX
SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.3e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMGAQAFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMGAQAFHQGESSAAAFQAAHARFVA 60

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97

RESULT 5
AAB35221
ID AAB35221 standard; protein; 97 AA.
XX
AC AAB35221;
DT 24-APR-2001 (first entry)
DE M tuberculosis Rv0287 protein.
XX
KW Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c; Rv1037c;
KW Rv2346c; Rv2348c; Rv2853c; Rv2654c; Rv3020c; Rv3444c; Rv3445c; Rv3890c;
KW Rv3891c; Rv3904c; Rv3905c.
XX
OS Mycobacterium tuberculosis.
XX
XX WO200104151-A2.
XX
XX 18-JAN-2001.
XX
XX 13-JUL-2000; 2000WO-DX000398.
XX
XX 13-JUL-1999; 99DK-00001020.
XX 15-JUL-1999; 99US-0144011P.
XX (STAT-) STATENS SERUM INST.
XX
XX Andersen P, Skjot R;
PI
XX
XX WPI; 2001-091923/10.
XX N-PSDB; AAF24411.
XX
XX New polypeptide encoded by a member of the esat-6 gene family for
PT immunizing against and diagnosis of tuberculosis.
XX
XX Example 2; Page 67; 80pp; English.
XX

The present invention provides the protein and coding sequences for
CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
CC proteins include Rv0287, Rv1036c, Rv1037c, Rv2346c, Rv2348c, Rv2653c,
CC Rv2654c, Rv3020c, Rv3444c, Rv3445c, Rv3890c, Rv3891c, Rv3904c and
CC Rv3905c. These can be used to produce vaccines against, and in the
CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
CC the proteins of the invention
XX
SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.3e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 Db 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97
 Db 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97

RESULT 6
 ID AAU08229 standard; protein; 97 AA.
 AC AAU08229;
 XX
 DT 17-DEC-2001 (first entry)
 DE Mycobacterium tuberculosis polypeptide for MTb98..
 XX
 DE Mycobacterium tuberculosis polypeptide for MTb98..
 XX
 KW Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
 KW immunostimulant; MTb98.
 XX
 OS Mycobacterium tuberculosis.
 XX
 OS WO200162893-A2.
 XX
 PN 30-AUG-2001.
 XX
 PD 26-FEB-2001; 2001WO-US005992.
 XX
 PF 25-FEB-2000; 2000US-0185037P.
 XX
 PR 08-AUG-2000; 2000US-0223828P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Skeiky Y, Ovendale P, Jen S, Lodes M;
 XX
 DR WPI; 2001-536638/59.
 XX
 DR An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
 PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
 PT response to and inhibiting development of a Mycobacterium infection.
 XX
 PS Disclosure; Page 150; 161pp; English.
 XX
 CC The present invention relates to the isolation of Mycobacterium
 CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
 CC encoding them. The invention describes compounds and methods for the
 CC diagnosis of tuberculosis or for inducing protective immunity against
 CC tuberculosis. The compounds comprise at least one immunogenic portion of
 CC one or more Mycobacterium proteins and nucleic acid molecules encoding
 CC such polypeptides. The Mycobacterium proteins and nucleic acid molecules
 CC of the invention and antibodies directed against the Mycobacterium
 CC proteins may be used in vaccines for immunisation against Mycobacterium
 CC infections. The nucleic acids encoding the Mycobacterium proteins may be
 CC used in gene therapy. The present sequence represents the M. tuberculosis
 CC polypeptide for MTb9.8
 XX
 SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 Db 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97
 Db 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97

Db 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97

RESULT 7
 ID AAU08209 standard; protein; 97 AA.
 AC AAU08209;
 XX
 DT 17-DEC-2001 (first entry)
 DE Mycobacterium tuberculosis antigen Tb472.
 XX
 DE Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
 KW immunostimulant; antigen Tb472.
 XX
 OS Mycobacterium tuberculosis.
 XX
 OS Key Location/Qualifiers
 FT Misc-difference 89.97
 FT /note= "Encoded by G"
 XX
 PN WO200162893-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 26-FEB-2001; 2001WO-US005992.
 XX
 PR 25-FEB-2000; 2000US-0185037P.
 XX
 PR 08-AUG-2000; 2000US-0223828P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Skeiky Y, Ovendale P, Jen S, Lodes M;
 XX
 DR WPI; 2001-536638/59.
 XX
 DR N-PSDB; AAS12467.
 XX
 DR An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
 PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
 PT response to and inhibiting development of a Mycobacterium infection.
 XX
 PS Example 1; Page 145-146; 161pp; English.
 XX
 CC The present invention relates to the isolation of Mycobacterium
 CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
 CC encoding them. The invention describes compounds and methods for the
 CC diagnosis of tuberculosis or for inducing protective immunity against
 CC tuberculosis. The compounds comprise at least one immunogenic portion of
 CC one or more Mycobacterium proteins and nucleic acid molecules encoding
 CC such polypeptides. The Mycobacterium proteins and nucleic acid molecules
 CC of the invention and antibodies directed against the Mycobacterium
 CC proteins may be used in vaccines for immunisation against Mycobacterium
 CC infections. The nucleic acids encoding the Mycobacterium proteins may be
 CC used in gene therapy. The present sequence represents M. tuberculosis
 CC antigen Tb472
 XX
 SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 Db 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97
 Db 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97

CC	heterologous antigens, and a variety of vaccines to prophylactically immunise mammals
CC	fusion proteins are used as vaccines to prophylactically immunise mammals
CC	(especially humans) against infection by Mycobacteria. The compositions
CC	contain at least 2 heterologous antigens that increase the serological
CC	sensitivity of individuals infected with tuberculosis, a disease
CC	frequently affecting patients with acquired immunodeficiency disease,
CC	AIDS
XX	
XX	Sequence 97 AA;
XX	
XX	Query Match 100.0%; Score 468; DB 4; Length 97;
XX	Best Local Similarity 100.0%; Pred. No. 2.3e-48;
XX	Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	
QY	1 MSLLDARIHPOLVASQSAFAAKAGLMRHTIGQAEQAMSAQAQAFHOGESSAAFOAAHAFVA 60
DB	1 MSLLDARIHPOLVASQSAFAAKAGLMRHTIGQAEQAMSAQAQAFHOGESSAAFOAAHAFVA 60
QY	61 AAAYNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
DB	61 AAAYNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
XX	
XX	RESULT 9
XX	AAE29712
XX	ID AAE29712 standard; protein; 97 AA.
XX	AC AAE29712;
XX	DT 27-JAN-2003 (first entry)
XX	DE Mycobacterium tuberculosis MSL antigenic protein.
XX	KW Vaccine; immunity; diagnostic agent; gene therapy; MSL antigen; MTB9.8.
XX	OS Mycobacterium tuberculosis.
XX	FH Key Location/Qualifiers
XX	FT Misc-difference 98..97 /note= "Encoded by GCTG"
XX	XX WO200272792-A2.
XX	XX 19-SEP-2002.
XX	XX 13-MAR-2002; 2002WO-US008223.
XX	XX 13-MAR-2001; 2001US-0275837P.
XX	XX (CORI-) CORIXA CORP.
XX	XX Skeiky Y, Brannon M, Guderian J;
XX	XX WPI; 2002-759844/82.
XX	XX N-PSDB; AAD47088.
XX	XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeifF,
XX	XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX	XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX	XX tuberculosis.
XX	XX Disclosure; Page 101; 155pp; English.
XX	XX The invention relates to a recombinant nucleic acid molecule encoding a
XX	XX fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX	XX polynucleotide sequence encoding an antigen or an antigenic fragment from
XX	XX Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX	XX polypeptide or its fragment. The Leishmania polynucleotide is selected
XX	XX from TSA, LeifF, M15, and 6H polynucleotides. Sequences of the invention
XX	XX are used in methods for eliciting immune response in mammals. They are
XX	XX useful as vaccines to elicit protective immunity against pathogenic
XX	XX microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX	XX polypeptides are used for enhancing the expression of polynucleotides, as
XX	XX polypeptides are used for enhancing the expression of polynucleotides, as

CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis MSL antigenic protein. MSL is also referred to as MTB9.8
 XX
 SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 5; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLDDAHIPOLVASQSAFAKAGLMRHTTIGQAFQQAAMSAQAFHQGESSAAFAQAAHARFVA 60
 Db 1 MSLDDAHIPOLVASQSAFAKAGLMRHTTIGQAFQQAAMSAQAFHQGESSAAFAQAAHARFVA 60

Qy 61 AAANKVNTLLDVAQANLGEAAGTVAADAAAASVTYGF 97
 Db 61 AAANKVNTLLDVAQANLGEAAGTVAADAAAASVTYGF 97

RESULT 10
 AAEL17576
 ID AAEL17576 standard; protein; 97 AA.
 XX
 AC AAEL17576;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis MTB9.8 (MSL) protein.
 XX
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB9.8; MSL protein.
 XX
 OS Mycobacterium tuberculosis.
 OS
 FH Key Location/Qualifiers
 FH Misc-difference 98..97 /note= "Encoded by GCTG"
 FT
 FT
 XX WO200198460-A2.
 XX
 XX 27-DEC-2001.
 PD
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 PF
 XX
 XX 20-JUN-2000; 2000US-00597796.
 PR
 XX 01-FEB-2001; 2001US-0265737P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 XX
 PI Skeiky Y, Reed S, Alderson M;
 XX
 DR WPI; 2002-147798/19.
 DR N-PSDB; AAD28346.
 XX
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 9; Page 117; 136pp; English.
 XX
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the

CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is M. tuberculosis MTB9.8
 CC (MSL) protein
 XX
 SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 5; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLDDAHIPOLVASQSAFAKAGLMRHTTIGQAFQQAAMSAQAFHQGESSAAFAQAAHARFVA 60
 Db 1 MSLDDAHIPOLVASQSAFAKAGLMRHTTIGQAFQQAAMSAQAFHQGESSAAFAQAAHARFVA 60

Qy 61 AAANKVNTLLDVAQANLGEAAGTVAADAAAASVTYGF 97
 Db 61 AAANKVNTLLDVAQANLGEAAGTVAADAAAASVTYGF 97

RESULT 11
 AAY32067
 ID AAY32067 standard; protein; 299 AA.
 XX
 AC AAY32067;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen fusion protein Mtb31f.
 XX
 KW Tuberculosis; antigen; fusion protein; Mtb31f; DPV; MTL; MSL; diagnosis;
 KW therapy; vaccine; immunogen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FH Peptide 1..8 /note= "Met/His tag"
 FT Protein 9..90 /note= "DPV"
 FT Protein 93..186 /note= "MTL"
 FT Protein 189..285 /note= "MSL"
 FT
 XX WO9951748-A2.
 XX
 XX 14-OCT-1999.
 PD
 XX
 XX 07-APR-1999; 99WO-US007717.
 PF
 XX 30-APR-1998; 98US-00065556.
 PR 07-DEC-1998; 98US-00223040.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky YAW, Alderson M, Campos-Neto A;
 PI
 XX WPI; 1999-601610/51.
 DR N-PSDB; AAZ20202.
 XX
 XX New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.
 XX
 XX Claim 1; Fig 9A-B; 83pp; English.
 PS
 XX This sequence represents a recombinant Mycobacterium tuberculosis tri-
 CC antigen fusion protein, termed Mtb31f, composed of the antigens DPV, MTL
 CC and MSL. The fusion protein is expressed in host cells using a vector
 CC carrying a polynucleotide (see AAZ20202) comprising the coding sequences

CC for the 3 antigens. The invention provides fusion proteins (see AAY32059-
CC 71) containing at least 2 M. tuberculosis antigens. The new fusion
CC proteins and polynucleotides encoding them are useful as vaccines for
CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
CC monitoring of disease progression, and treatment of tuberculosis. They
CC are more effective immunogens than mixtures of the individual protein
CC components
XX
XX Sequence 299 AA;

Query Match 100.0%; Score 468; DB 2; Length 299;
Best Local Similarity 100.0%; Pred. No. 9.6e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAMSAQAAPHQGESAAFOAAHARFVA 60
Db 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAMSAQAAPHQGESAAFOAAHARFVA 248
QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
Db 249 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 285

RESULT 12
AAE29723
ID AAE29723 standard; protein; 299 AA.
XX
AC AAE29723;
XX
DT 29-AUG-2003 (revised)
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium sp. DPV-MTI-MSL-MTCC#2 fusion protein.
XX
KW Vaccine; immunity; diagnostic agent; gene therapy; DPV-MTI-MSL-MTCC#2;
KW MTB71F; MTB8.4; MTB9.9A; MTB9.8; MTB41; fusion protein.
XX
OS Mycobacterium tuberculosis.
OS Mycobacterium sp.
OS Chimeric.
XX
PN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WPI; 2002-759844/82.
DR N-PSDB; AAD47100.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
PT tuberculosis.
XX
PS Disclosure; Page 114-115; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion

CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC MTB71F fusion protein. This fusion protein comprises DPV (MTB8.4)-MTI
CC (MTB9.9A)-MSL (MTB9.8) protein sequences from M. tuberculosis and MTCC#2
CC (MTB41) protein from Mycobacterium sp. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
XX Sequence 299 AA;

Query Match 100.0%; Score 468; DB 5; Length 299;
Best Local Similarity 100.0%; Pred. No. 9.6e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAMSAQAAPHQGESAAFOAAHARFVA 60
Db 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAMSAQAAPHQGESAAFOAAHARFVA 248
QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
Db 249 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 285

RESULT 13
AAE17587
ID AAE17587 standard; protein; 299 AA.
XX
AC AAE17587;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species MTB71F fusion protein.
XX
KW Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB71F; DPV-MTI-MSL-MTCC#2 protein.
XX
OS Mycobacterium sp.
XX
PN WO200198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US019959.
XX
PR 20-JUN-2000; 2000US-00597796.
PR 01-FEB-2001; 2001US-0265737P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Alderson M;
XX
DR WPI; 2002-147798/19.
DR N-ESDB; AAD28358.
XX

XX Composition comprising MTB39 antigen and MTB32A antigen from
XX Mycobacterium species, useful for eliciting immune response in a subject.
XX Disclosure; Page 130-131; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the

CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB71F (DPV-MTI-MSL-MTCC#2) fusion protein
 XX
 SQ Sequence 299 AA;

Query Match 100.0%; Score 468; DB 5; Length 299;
 Best Local Similarity 100.0%; Pred. No. 9.6e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAPAAKAGLMRHTIGQAEQAMSAQAFHQGESSAAFOAAHARFVA 60
 Db 189 MSLLDAHIPOLVASQSAPAAKAGLMRHTIGQAEQAMSAQAFHQGESSAAFOAAHARFVA 248
 QY 61 AAANKVNTLLDVAQANLGEAAAGTYVAADAAASITYTGF 97
 Db 249 AAANKVNTLLDVAQANLGEAAAGTYVAADAAASITYTGF 285

RESULT 14
 AAU74596
 ID AAU74596 standard; protein; 306 AA.

AC AAU74596;
 XX
 DT 29-AUG-2003 (revised)
 DT 08-MAY-2002 (first entry)
 XX
 XX Antigenic fusion protein DPV-MTI-MSL (Mtb31f).
 XX
 KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; DPV-MTI-MSL; Mtb31f.
 XX
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 300
 FT /label= OTHER
 FT /note= "OTHER= Xaa. Xaa= In frame stop codon"
 XX
 PN US2002009459-A1.
 XX
 PD 24-JAN-2002.
 XX
 PF 07-APR-1999; 99US-00287849.
 XX
 PR 13-MAR-1997; 97US-00818112.
 PR 01-OCT-1997; 97US-00942578.
 PR 18-FEB-1998; 98US-00025197.
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 XX (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.

PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 XX
 DR WPI; 2002-171134/22.
 DR N-PSDB; ABK14136.

XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis.
 XX
 PS Claim 1; Fig 9; 62pp; English.

XX The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention. (Updated on 29-AUG-2003 to standardise OS field)
 XX
 SQ Sequence 306 AA;

Query Match 100.0%; Score 468; DB 5; Length 306;
 Best Local Similarity 100.0%; Pred. No. 9.9e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAPAAKAGLMRHTIGQAEQAMSAQAFHQGESSAAFOAAHARFVA 60
 Db 189 MSLLDAHIPOLVASQSAPAAKAGLMRHTIGQAEQAMSAQAFHQGESSAAFOAAHARFVA 248
 QY 61 AAANKVNTLLDVAQANLGEAAAGTYVAADAAASITYTGF 97
 Db 249 AAANKVNTLLDVAQANLGEAAAGTYVAADAAASITYTGF 285

RESULT 15
 AAY32065
 ID AAY32065 standard; protein; 433 AA.

AC AAY32065;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen fusion protein Mtb46f.
 XX
 KW Tuberculosis; antigen; fusion protein; Mtb46f; ERD14; DPV; MTI; MSL;
 KW diagnosis; therapy; vaccine; immunogen.

OS Mycobacterium tuberculosis.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..8
 FT /note= "Met/His tag"
 FT Protein 9..152
 FT /note= "ERD14"
 FT Protein 155..236
 FT /note= "DPV"
 FT Protein 239..332
 FT /note= "MTI"
 FT Protein 335..433
 FT /note= "MSL"

WO9951748-A2.

PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US007717.
 XX
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.

XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky YAW, Alderson M, Campos-Neto A;
 PI
 DR WPI; 1999-601610/51.
 DR N-PSDB; AAZ20200.
 XX
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.

PS Claim 1; Fig 7A-B; 83pp; English.
XX This sequence represents a recombinant Mycobacterium tuberculosis tetra-
CC antigen fusion protein, termed Mt46f, composed of the antigens ERD14,
CC DPV, MTI and MSL. The fusion protein is expressed in host cells using a
CC vector carrying a polynucleotide (see AAZ20200) comprising the coding
CC sequences for the 4 antigens. The invention provides fusion proteins (see
CC AAY32059-71) containing at least 2 M. tuberculosis antigens. The new
CC fusion proteins and polynucleotides encoding them are useful as vaccines
CC for preventing tuberculosis (claimed), for diagnosis (via in vitro assays
CC or intradermal skin tests for detection of anti-M. tuberculosis
CC antibodies), monitoring of disease progression, and treatment of
CC tuberculosis. They are more effective immunogens than mixtures of the
CC individual protein components
XX
SQ Sequence 433 AA;

Query Match 100.0%; Score 468; DB 2; Length 433;
Best Local Similarity 100.0%; Pred. No. 1.5e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESAAFOAHAREVA 60
DB 335 MSLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESAAFOAHAREVA 394
QY 61 AAKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97
DB 395 AAKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 431

Search completed: June 23, 2004, 17:12:30
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 23 Seconds
(without alignments)
217.727 Million cell updates/sec

Title: US-09-886-349A-24
Perfect score: 468
Sequence: 1 MSLDHAHPQLVASQSAFAA.....EAGTYVADAAASATYTG 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	100.0	97	4	US-09-073-009-109
2	468	100.0	97	4	US-09-073-009-143
3	468	100.0	97	4	US-09-073-010-109
4	468	100.0	97	4	US-09-073-010-143
5	468	100.0	299	4	US-09-287-849-19
6	468	100.0	433	4	US-09-287-849-14
7	468	100.0	710	4	US-09-287-849-16
8	468	100.0	856	4	US-09-287-849-12
9	87.5	18.7	95	3	US-08-818-112-88
10	87.5	18.7	95	4	US-08-818-111-89
11	87.5	18.7	95	4	US-09-056-556-88
12	87.5	18.7	95	4	US-09-072-596-89
13	87.5	18.7	95	4	US-09-072-967-88
14	87.5	18.7	100	3	US-08-818-112-115
15	87.5	18.7	100	4	US-08-818-111-110
16	87.5	18.7	100	4	US-09-056-556-115
17	87.5	18.7	100	4	US-09-116-492A-5
18	87.5	18.7	100	4	US-09-072-596-110
19	87.5	18.7	100	4	US-09-072-967-115
20	87.5	18.7	358	4	US-09-287-849-8
21	87.5	18.7	802	4	US-09-056-556-214
22	87.5	18.7	802	4	US-09-072-596-209
23	87.5	18.7	802	4	US-09-072-596-346
24	87.5	18.7	802	4	US-09-072-967-214
25	87.5	18.7	802	4	US-09-072-967-351
26	87.5	18.7	802	4	US-09-287-849-10
27	86.5	18.5	80	3	US-08-818-112-117

28	86.5	18.5	80	4	US-08-818-111-112	Sequence 112, App
29	86.5	18.5	80	4	US-09-056-556-117	Sequence 117, App
30	86.5	18.5	80	4	US-09-072-596-112	Sequence 112, App
31	86.5	18.5	80	4	US-09-072-967-117	Sequence 117, App
32	86	18.4	18	4	US-09-073-009-124	Sequence 124, App
33	86	18.4	18	4	US-09-073-010-124	Sequence 124, App
34	79.5	17.0	92	4	US-09-344-529-2	Sequence 2, Appli
35	79.5	17.0	361	4	US-09-543-681A-5390	Sequence 5390, Ap
36	77.5	16.6	400	4	US-09-073-009-126	Sequence 126, App
37	77.5	16.6	400	4	US-09-073-010-126	Sequence 126, App
38	77	16.5	15	4	US-09-073-009-118	Sequence 118, App
39	77	16.5	15	4	US-09-073-010-118	Sequence 118, App
40	76.5	16.3	286	4	US-09-489-039A-12707	Sequence 12707, A
41	75.5	16.1	469	4	US-09-489-039A-13565	Sequence 13565, A
42	75	16.0	15	4	US-09-073-009-114	Sequence 114, App
43	75	16.0	15	4	US-09-073-010-114	Sequence 114, App
44	75	16.0	339	1	US-08-433-854-4	Sequence 4, Appli
45	75	16.0	339	1	US-08-174-745A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-073-009-109
; Sequence 109, Application US/09073009
; Patent No. 6555653
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-073-009-109

TUBERCULOSIS AND MEI

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.7e-53;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSLDAHPQLVASQSAFAAKAGLMRHTTIGQAQAMSAQAFAHOGESSAAFAAHARFVA	60
DB	1	MSLDAHPQLVASQSAFAAKAGLMRHTTIGQAQAMSAQAFAHOGESSAAFAAHARFVA	60

QY 61 AAANKVNTLLDVAQNLGEAAGTYVAADAAAATYTG 97
 DB 61 AAANKVNTLLDVAQNLGEAAGTYVAADAAAATYTG 97

RESULT 2

US-09-073-009-143
 ; Sequence 143, Application US/09073009
 ; Patent No. 6555653
 ; GENERAL INFORMATION:
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Campos-Neto, Antonio
 ; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
 ; NUMBER OF SEQUENCES: 144
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY
 ; STREET: 6300 Columbia Center, 701 Fifth Ave.
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: US
 ; ZIP: 98104

TUBERCULOSIS AND ME

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/073,009
 FILING DATE: 05-MAY-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.441C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 206-622-4900
 TELEFAX: 206-682-6031
 INFORMATION FOR SEQ ID NO: 143:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 97 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-09-073-009-143
 Query Match 100.0%; Score 468; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 8.7e-53;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60
 DB 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60
 QY 61 AAANKVNTLLDVAQNLGEAAGTYVAADAAAATYTG 97
 DB 61 AAANKVNTLLDVAQNLGEAAGTYVAADAAAATYTG 97

US-09-073-010-143
 ; Sequence 143, Application US/09073010
 ; Patent No. 6613881
 ; GENERAL INFORMATION:
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Campos-Neto, Antonio
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
 ; NUMBER OF SEQUENCES: 144
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY
 ; STREET: 6300 Columbia Center, 701 Fifth Ave.
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: US
 ; ZIP: 98104

RESULT 3

US-09-073-010-109
 ; Sequence 109, Application US/09073010
 ; Patent No. 6613881
 ; GENERAL INFORMATION:
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Campos-Neto, Antonio
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
 ; NUMBER OF SEQUENCES: 144

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY
 STREET: 6300 Columbia Center, 701 Fifth Ave.
 CITY: Seattle
 STATE: Washington
 COUNTRY: US
 ZIP: 98104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/073,010
 FILING DATE: 05-MAY-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.440C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 206-622-4900
 TELEFAX: 206-682-6031
 INFORMATION FOR SEQ ID NO: 109:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 97 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-09-073-010-109
 Query Match 100.0%; Score 468; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 8.7e-53;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60
 DB 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60
 QY 61 AAANKVNTLLDVAQNLGEAAGTYVAADAAAATYTG 97
 DB 61 AAANKVNTLLDVAQNLGEAAGTYVAADAAAATYTG 97

US-09-073-010-143
 ; Sequence 143, Application US/09073010
 ; Patent No. 6613881
 ; GENERAL INFORMATION:
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Campos-Neto, Antonio
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
 ; NUMBER OF SEQUENCES: 144
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY
 ; STREET: 6300 Columbia Center, 701 Fifth Ave.
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: US
 ; ZIP: 98104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/073,010
 FILING DATE: 05-MAY-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:


```
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.440C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-073-010-143

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.7e-53;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFOAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFOAAHARFVA 60
QY 61 AAANKVNTLLDVAQANLGEAACTVVAADAAAASYTGF 97
DB 61 AAANKVNTLLDVAQANLGEAACTVVAADAAAASYTGF 97

RESULT 5
US-09-287-849-19
; Sequence 19, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-19

Query Match 100.0%; Score 468; DB 4; Length 299;
Best Local Similarity 100.0%; Pred. No. 3.8e-52;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFOAAHARFVA 60
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFOAAHARFVA 248
QY 61 AAANKVNTLLDVAQANLGEAACTVVAADAAAASYTGF 97

; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.440C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-073-010-143

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.7e-53;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFOAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFOAAHARFVA 60
QY 61 AAANKVNTLLDVAQANLGEAACTVVAADAAAASYTGF 97
DB 61 AAANKVNTLLDVAQANLGEAACTVVAADAAAASYTGF 97

RESULT 6
US-09-287-849-14
; Sequence 14, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-14

Query Match 100.0%; Score 468; DB 4; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.2e-52;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFOAAHARFVA 60
DB 335 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFOAAHARFVA 394
QY 61 AAANKVNTLLDVAQANLGEAACTVVAADAAAASYTGF 97
DB 395 AAANKVNTLLDVAQANLGEAACTVVAADAAAASYTGF 431

RESULT 7
US-09-287-849-16
; Sequence 16, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-16
```

PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 16
LENGTH: 710
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-16

Query Match 100.0%; Score 468; DB 4; Length 710;
Best Local Similarity 100.0%; Pred. No. 1.2e-51;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAQAAMSAQAAPHQGSAAFAQAARFVA 60
Db 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAQAAMSAQAAPHQGSAAFAQAARFVA 248

Qy 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
Db 249 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 285

RESULT 8
US-09-287-849-12
Sequence 12, Application US/09287849
Patent No. 6627198
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287,849
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 12
LENGTH: 856
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:penta-fusion
US-09-287-849-12

Query Match 100.0%; Score 468; DB 4; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.5e-51;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAQAAMSAQAAPHQGSAAFAQAARFVA 60
Db 335 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAQAAMSAQAAPHQGSAAFAQAARFVA 394

Qy 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
Db 395 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 431

RESULT 9
US-08-818-112-88
Sequence 88, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-88

Query Match 18.7%; Score 87.5; DB 3; Length 95;
Best Local Similarity 34.9%; Pred. No. 0.00071;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

Qy 18 FAKAGLMRHTTIGQAQAAMSAQAAPHQGSAAFAQAARFVAQAANKVNTLLDVAQANL 77
Db 13 PERISGDLTKIQDOVESTAGSLQGWGGAAGTAQAQAAVVFQEAANKQKQELDEISTNIR 72

Qy 78 EAAGTYVAAD---AAAATYTG 97
Db 73 QAGVYGRADEEQQAALSSQMGF 95

RESULT 10
US-08-818-111-89
Sequence 89, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio

```
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-111-89
;
; Query Match 18.7%; Score 87.5; DB 4; Length 95;
; Best Local Similarity 34.9%; Pred.No.0.00071;
; Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;
;
; QY 18 FAKAGLMRHTIGQAEQAAWSAQAFHQGESSAAFOAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWRGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; RESULT 11
; US-09-056-556-88
; Sequence 88, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; APPLICATION NUMBER: US/09/056,556
;
; QY 18 FAKAGLMRHTIGQAEQAAWSAQAFHQGESSAAFOAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWRGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; RESULT 12
; US-09-072-596-89
; Sequence 89, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; APPLICATION NUMBER: US/09/072,596
;
; QY 18 FAKAGLMRHTIGQAEQAAWSAQAFHQGESSAAFOAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWRGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; Query Match 18.7%; Score 87.5; DB 4; Length 95;
; Best Local Similarity 34.9%; Pred.No.0.00071;
; Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;
;
; QY 18 FAKAGLMRHTIGQAEQAAWSAQAFHQGESSAAFOAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWRGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; RESULT 12
; US-09-072-596-89
; Sequence 89, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; APPLICATION NUMBER: US/09/072,596
;
; QY 18 FAKAGLMRHTIGQAEQAAWSAQAFHQGESSAAFOAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWRGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; Query Match 18.7%; Score 87.5; DB 4; Length 95;
; Best Local Similarity 34.9%; Pred.No.0.00071;
; Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;
;
; QY 18 FAKAGLMRHTIGQAEQAAWSAQAFHQGESSAAFOAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWRGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; RESULT 11
; US-09-056-556-88
; Sequence 88, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND TREATM
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
```

STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-89

Query Match 18.7%; Score 87.5; DB 4; Length 95;
Best Local Similarity 34.9%; Pred. No. 0.00071;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

QY 18 FAKAGLMRHTIGAEQAAMSAQAFHQGESSAAFAQAAHARFVAAAKVNTLLDVAQANLG 77

Db 13 FERISGDLTKTQIDOVSTAGSLQGQWRGAAGTAAQAAVVRFQEAANKQKQELDEISTNIR 72

QY 78 EAAGTYVAAD---AAASTYTG 97

Db 73 QAGVYSRADEEQQALSSQMGF 95

RESULT 13

US-09-072-967-88

; Sequence 88, Application US/09072967

; Patent No. 6592877

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 355

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,967

; FILING DATE: 05-MAY-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.411C9

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 88:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 95 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-072-967-88

Query Match 18.7%; Score 87.5; DB 4; Length 95;

Best Local Similarity 34.9%; Pred. No. 0.00071;

Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

QY 18 FAKAGLMRHTIGAEQAAMSAQAFHQGESSAAFAQAAHARFVAAAKVNTLLDVAQANLG 77

Db 13 FERISGDLTKTQIDOVSTAGSLQGQWRGAAGTAAQAAVVRFQEAANKQKQELDEISTNIR 72

QY 78 EAAGTYVAAD---AAASTYTG 97

Db 73 QAGVYSRADEEQQALSSQMGF 95

RESULT 14

US-08-818-112-115

; Sequence 115, Application US/08818112

; Patent No. 6290369

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 153

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/818,112

; FILING DATE: 13-MAR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.411C6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 115:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 100 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-08-818-112-115

Query Match 18.7%; Score 87.5; DB 3; Length 100;

Best Local Similarity 34.9%; Pred. No. 0.00076;

Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

QY 18 FAKAGLMRHTIGAEQAAMSAQAFHQGESSAAFAQAAHARFVAAAKVNTLLDVAQANLG 77

Db 18 FERISGDLTKTQIDOVSTAGSLQGQWRGAAGTAAQAAVVRFQEAANKQKQELDEISTNIR 77

QY 78 EAAGTYVAAD---AAASTYTG 97

Db 78 QAGVYSRADEEQQALSSQMGF 100

RESULT 15

US-08-818-111-110

; Sequence 110, Application US/08818111

; Patent No. 6338852

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-111-110

Query Match 18.7%; Score 87.5; DB 4; Length 100;
Best Local Similarity 34.9%; Pred. No. 0.00076;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;
QY 18 EAAGTYVAAD--AAASTYTG 97
Db 18 FERISGLTKQIDQVESTAGSLGGQWRGAAGTAQAQAAVVRFOEAANKKQKQELDEISTNIR 77
QY 78 EAAGTYVAAD--AAASTYTG 97
Db 78 QAGVQYSRADERQQQALSSQMGF 100

Search completed: June 23, 2004, 17:10:43
Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:12:36 ; Search time 43 Seconds
(without alignments)
636.847 Million cell updates/sec

Title: US-09-886-349A-24

Perfect score: 468
Sequence: 1 MSLDHPQLVQSAPFAA.....EAACTYVAADAAASTVTGF 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	100.0	97	US-09-073-009-109	Sequence 109, App
2	468	100.0	97	US-09-073-009-143	Sequence 143, App
3	468	100.0	97	US-09-793-306-109	Sequence 109, App
4	468	100.0	97	US-09-793-306-143	Sequence 143, App
5	468	100.0	97	US-09-886-349A-24	Sequence 24, Appl
6	468	100.0	97	US-10-098-732A-24	Sequence 24, Appl
7	468	100.0	299	US-09-287-849-19	Sequence 19, Appl
8	468	100.0	299	US-09-886-349A-47	Sequence 47, Appl
9	468	100.0	299	US-10-359-460-19	Sequence 19, Appl
10	468	100.0	299	US-10-098-732A-47	Sequence 47, Appl
11	468	100.0	433	US-09-287-849-14	Sequence 14, Appl
12	468	100.0	433	US-10-359-460-14	Sequence 14, Appl
13	468	100.0	710	US-09-287-849-16	Sequence 16, Appl
14	468	100.0	710	US-09-886-349A-49	Sequence 49, Appl
15	468	100.0	710	US-10-359-460-16	Sequence 16, Appl

Sequence 49, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 300, Appl
Sequence 37, Appl
Sequence 37, Appl
Sequence 6207, A
Sequence 61942, A
Sequence 35, Appl
Sequence 89, Appl
Sequence 88, Appl
Sequence 35, Appl
Sequence 5, Appl
Sequence 639, App
Sequence 110, App
Sequence 115, App
Sequence 8, Appl
Sequence 8, Appl
Sequence 10, Appl
Sequence 209, App
Sequence 346, App
Sequence 314, App
Sequence 351, App
Sequence 10, Appl
Sequence 112, App
Sequence 117, App
Sequence 124, App
Sequence 124, App
Sequence 64750, A
Sequence 26, Appl

710 14 US-10-098-732A-49
856 9 US-09-287-849-12
856 14 US-10-359-460-12
856 14 US-10-080-170-300
352 75.2 98 14 US-10-393-449-37
104 12 US-10-177-725-37
104 14 US-10-177-725-37
90.5 19.3 184 12 US-10-282-122A-62027
21 30.5 19.3 405 12 US-10-282-122A-61942
22 89.5 19.1 163 12 US-09-886-349A-35
23 88.5 18.9 95 12 US-10-193-002-89
24 87.5 18.7 95 14 US-10-084-843-88
25 87.5 18.7 95 14 US-10-098-732A-35
26 87.5 18.7 95 14 US-10-140-045-5
27 87.5 18.7 100 14 US-10-080-170-639
28 87.5 18.7 100 14 US-10-193-002-115
29 87.5 18.7 100 14 US-10-084-843-115
30 87.5 18.7 358 9 US-09-287-849-8
31 87.5 18.7 358 14 US-10-359-460-8
32 87.5 18.7 802 9 US-09-287-849-10
33 87.5 18.7 802 14 US-10-193-002-209
34 87.5 18.7 802 14 US-10-193-002-346
35 87.5 18.7 802 14 US-10-084-843-214
36 87.5 18.7 802 14 US-10-084-843-351
37 87.5 18.7 802 14 US-10-359-460-10
38 87.5 18.7 802 14 US-10-193-002-112
39 87.5 18.5 80 14 US-10-084-843-117
40 86.5 18.5 80 14 US-10-084-843-117
41 86.5 18.5 80 14 US-10-084-843-117
42 86 18.4 18 9 US-09-073-009-124
43 86 18.4 18 9 US-09-793-306-124
44 86 18.4 461 12 US-10-282-122A-64750
45 85.5 18.3 104 12 US-10-393-449-26

ALIGNMENTS

RESULT 1

US-09-073-009-109
; Sequence 109, Application US/09073009
; Patent No. US20010012888A1
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104

TUBERCULOSIS AND MET

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/073.009
FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-073-009-109

Query Match 100.0%; Score 468; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLLDAHIPOLVASQSAFAKAGLMRHTTICQAEQAAMSAQAAPHQGESAAFOAAHARFVA 60

Db 1 MSLLDAHIPOLVASQSAFAKAGLMRHTTICQAEQAAMSAQAAPHQGESAAFOAAHARFVA 60

Qy 61 AAQKNTLLDVAQNLGEAAGTVAADAAAATYTG 97

Db 61 AAQKNTLLDVAQNLGEAAGTVAADAAAATYTG 97

RESULT 2

US-09-073-009-143

; Sequence 143, Application US/09073009

; Patent No. US2001001288A1

; GENERAL INFORMATION:

; APPLICANT: Alderson, Mark

; APPLICANT: Dillon, David C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Campos-Neto, Antonio

; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF

; NUMBER OF SEQUENCES: 144

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY

; STREET: 6300 Columbia Center, 701 Fifth Ave.

; CITY: Seattle

; STATE: Washington

; COUNTRY: US

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/073,009

; FILING DATE: 05-MAY-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maxi, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.441C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-622-4900

; TELEFAX: 206-682-6031

; INFORMATION FOR SEQ ID NO: 143:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 97 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-073-009-143

Query Match 100.0%; Score 468; DB 9; Length 97;

Best Local Similarity 100.0%; Pred. No. 4.4e-47;

Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLLDAHIPOLVASQSAFAKAGLMRHTTICQAEQAAMSAQAAPHQGESAAFOAAHARFVA 60

Db 1 MSLLDAHIPOLVASQSAFAKAGLMRHTTICQAEQAAMSAQAAPHQGESAAFOAAHARFVA 60

Qy 61 AAQKNTLLDVAQNLGEAAGTVAADAAAATYTG 97

Db 61 AAQKNTLLDVAQNLGEAAGTVAADAAAATYTG 97

TUBERCULOSIS AND ME

RESULT 3

US-09-793-306-109

; Sequence 109, Application US/09793306

; Patent No. US20020098200A1

; GENERAL INFORMATION:

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Skeiky, Yasir

; APPLICANT: Owendale, Pamela

; APPLICANT: Jen, Shyian

; APPLICANT: Lodes, Michael

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

; FILE REFERENCE: 014058-008740US

; CURRENT APPLICATION NUMBER: US/09/793,306

; CURRENT FILING DATE: 2001-02-26

; PRIOR APPLICATION NUMBER: US 60/185,037

; PRIOR FILING DATE: 2000-02-25

; PRIOR APPLICATION NUMBER: US 60/223,828

; PRIOR FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 164

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 109

; LENGTH: 97

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: Tb472 (MSL)

US-09-793-306-109

Query Match 100.0%; Score 468; DB 9; Length 97;

Best Local Similarity 100.0%; Pred. No. 4.4e-47;

Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLLDAHIPOLVASQSAFAKAGLMRHTTICQAEQAAMSAQAAPHQGESAAFOAAHARFVA 60

Db 1 MSLLDAHIPOLVASQSAFAKAGLMRHTTICQAEQAAMSAQAAPHQGESAAFOAAHARFVA 60

Qy 61 AAQKNTLLDVAQNLGEAAGTVAADAAAATYTG 97

Db 61 AAQKNTLLDVAQNLGEAAGTVAADAAAATYTG 97

RESULT 4

US-09-793-306-143

; Sequence 143, Application US/09793306

; Patent No. US20020098200A1

; GENERAL INFORMATION:

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Skeiky, Yasir

; APPLICANT: Owendale, Pamela

; APPLICANT: Jen, Shyian

; APPLICANT: Lodes, Michael

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

; FILE REFERENCE: 014058-008740US

; CURRENT APPLICATION NUMBER: US/09/793,306

; CURRENT FILING DATE: 2001-02-26

; PRIOR APPLICATION NUMBER: US 60/185,037

; PRIOR FILING DATE: 2000-02-25

; PRIOR APPLICATION NUMBER: US 60/223,828

; PRIOR FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 164

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 143

; LENGTH: 97

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: Mtb9.8

US-09-793-306-143


```
Query Match      100.0%; Score 468; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAAFQAAHARFVA 60
QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97

RESULT 5
US-09-886-349A-24
; Sequence 24, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB9.8 (MSL)
US-09-886-349A-24

Query Match      100.0%; Score 468; DB 12; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAAFQAAHARFVA 60
QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97

RESULT 6
US-10-098-732A-24
; Sequence 24, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
```

```
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB9.8 (MSL)
US-10-098-732A-24

Query Match      100.0%; Score 468; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAAFQAAHARFVA 60
QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97

RESULT 7
US-09-287-849-19
; Sequence 19, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 19
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-19

Query Match      100.0%; Score 468; DB 9; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAAFQAAHARFVA 60
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAAFQAAHARFVA 248
QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
DB 249 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 285

RESULT 8
US-09-886-349A-47
; Sequence 47, Application US/09886349A
; Publication No. US20040086523A1
```

GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Reed, Steven
APPLICANT: Alderson, Mark
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009070US
CURRENT APPLICATION NUMBER: US/09/886,349A
PRIOR FILING DATE: 2001-06-20
PRIOR FILING DATE: 2000-06-20
PRIOR FILING DATE: 2000-06-20
PRIOR FILING DATE: 2001-02-01
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 47
LENGTH: 299
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
OTHER INFORMATION: protein DPV-MTI-MSL (designated MTB31F) CDNA
US-09-886-349A-47

Query Match 100.0%; Score 468; DB 12; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAFFHQGESSAAFOAAHARFVA 60
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAFFHQGESSAAFOAAHARFVA 248
QY 61 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 97
DB 249 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 285

RESULT 9
US-10-359-460-19
Sequence 19, Application US/10359460
Publication No. US20030147911A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR FILING DATE: 2003-02-05
PRIOR FILING DATE: 1999-04-07
PRIOR FILING DATE: 1999-04-07
PRIOR FILING DATE: 1997-03-13
PRIOR FILING DATE: 1997-03-13
PRIOR FILING DATE: 1997-10-01
PRIOR FILING DATE: 1997-10-01
PRIOR FILING DATE: 1998-02-18
PRIOR FILING DATE: 1998-02-18
PRIOR FILING DATE: 1998-04-07
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 19
LENGTH: 299
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion

US-10-359-460-19
Query Match 100.0%; Score 468; DB 14; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAFFHQGESSAAFOAAHARFVA 60
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAFFHQGESSAAFOAAHARFVA 248
QY 61 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 97
DB 249 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 285

RESULT 10
US-10-098-732A-47
Sequence 47, Application US/10098732A
Publication No. US20030175294A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
TITLE OF INVENTION: Leishmania Antigen
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 60/275,837
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 47
LENGTH: 299
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
OTHER INFORMATION: protein DPV-MTI-MSL (designated MTB31F)
US-10-098-732A-47

Query Match 100.0%; Score 468; DB 14; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAFFHQGESSAAFOAAHARFVA 60
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAFFHQGESSAAFOAAHARFVA 248
QY 61 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 97
DB 249 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 285

RESULT 11
US-09-287-849-14
Sequence 14, Application US/09287849
Patent No. US20020009459A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287,849
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13


```

; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
; OTHER INFORMATION: protein DPV-MTI-MSL-WTCC#2 (designated MTb71F)
US-09-886-349A-49

```

```

Query Match      100.0%; Score 468; DB 12; Length 710;
Best Local Similarity 100.0%; Pred. No. 4.9e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLDDAHIPQIVASQSAFAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60
Db 189 MSLDDAHIPQIVASQSAFAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 248

QY 61 AAQKNTLLDVAQNLGEAAGTYVAADAAAATYTG 97
Db 249 AAQKNTLLDVAQNLGEAAGTYVAADAAAATYTG 285

```

```

RESULT 15
US-10-359-460-16
; Sequence 16, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Aldersen, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-10-359-460-16

```

```

Query Match      100.0%; Score 468; DB 14; Length 710;
Best Local Similarity 100.0%; Pred. No. 4.9e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLDDAHIPQIVASQSAFAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60
Db 189 MSLDDAHIPQIVASQSAFAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 248

QY 61 AAQKNTLLDVAQNLGEAAGTYVAADAAAATYTG 97

```

```

Db 249 AAQKNTLLDVAQNLGEAAGTYVAADAAAATYTG 285

Search completed: June 23, 2004, 17:18:23
Job time : 48 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 21 Seconds
(without alignments)
444.313 Million cell updates/sec

Title: US-09-886-349A-24
Perfect score: 468
Sequence: 1 MSLLDAHIPQLVASQSAFAA.....BAAGTYVAADAAASTYTGTF 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: Pirl:*
2: Pirl:*
3: Pirl:*
4: Pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	100.0	97	E70836	probable transcript
2	442	94.4	97	G70857	probable PE protei
3	356	76.1	98	A87226	PE-family protein
4	92	19.7	135	A70659	hypothetical prote
5	87.5	18.7	100	H70802	hypothetical prote
6	87	18.6	1156	T34852	probable secreted
7	86	18.4	331	F70820	hypothetical glyci
8	85	18.2	431	F70571	hypothetical glyci
9	85	18.2	1147	T35781	hypothetical prote
10	84	17.9	457	H70820	hypothetical glyci
11	80.5	17.2	361	G70682	hypothetical glyci
12	80.5	17.2	394	G70881	probable PPE prote
13	80	17.1	405	F82181	galactokinase VC15
14	77.5	16.6	393	G70929	probable PPE prote
15	77.5	16.6	394	A70504	probable PPE prote
16	77.5	16.6	1238	T03465	probable exonuclea
17	76.5	16.3	175	A70824	hypothetical glyci
18	76	16.2	376	AG0592	tolA protein limpo
19	76	16.2	891	JN0867	peroxinectin-like
20	76	16.2	894	PN0667	peroxinectin-like
21	75.5	16.1	205	S19114	cgr-1 protein - C
22	75.5	16.1	582	IKESCA	colicin A - Citrob
23	75.5	16.1	882	G70812	hypothetical glyci
24	75	16.0	99	G70560	probable PE protei
25	75	16.0	339	JT0756	group-V allergen 1
26	75	16.0	1137	G70868	probable regulator
27	74.5	15.9	436	G81655	conserved hypothet
28	74.5	15.9	498	G70720	hypothetical glyci
29	74.5	15.9	582	B70953	hypothetical glyci

30 74.5 15.9 741 2 G70917
31 74.5 15.9 1538 2 H70846
32 74 15.8 97 2 S02376
33 74 15.8 401 1 QXBpIL
34 74 15.8 978 2 C59237
35 73.5 15.7 102 2 C70836
36 73.5 15.7 558 2 E70756
37 73.5 15.7 646 2 T42396
38 73.5 15.7 810 2 B84185
39 73.5 15.7 2082 2 T37056
40 73 15.6 357 2 G70773
41 73 15.6 805 2 T25795
42 72.5 15.5 801 2 E70824
43 72.5 15.5 837 2 E70835
44 72.5 15.5 957 2 D70835
45 72.5 15.5 987 2 E70808

ALIGNMENTS

RESULT 1

E70836
probable transcription regulator Rv0287 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70836
R:Coate, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70836
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-97 <COL>
A:Cross-references: GB:AL021930; GB:AL123456; NID:G3261524; PIDN:CAA17362.1; PID:el25247
A:Experimental source: strain H37RV
C:Genetics:
C:Gene: Rv0287

Query Match 100.0%; Score 468; DB 2; Length 97;

Best Local Similarity 100.0%; Pred. No. 1.3e-38;

Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPQLVASQSAPAAKAGLMRHTIGQAEQAASQAQAFHOGESSAAFOAAHARVA 60

DB 1 MSLLDAHIPQLVASQSAPAAKAGLMRHTIGQAEQAASQAQAFHOGESSAAFOAAHARVA 60

QY 61 AAANKVTLDDVAQNLGEAGTYVAADAAASTYTGTF 97

DB 61 AAANKVTLDDVAQNLGEAGTYVAADAAASTYTGTF 97

RESULT 2

G70857

probable PE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70857

R:Coate, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70857

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-97 <COL>

A; Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAA16105.1; PID:ei23776
A; Experimental source: strain H37Rv
C; Genetics:
A; Gene: PE

Query Match 94.4%; Score 442; DB 2; Length 97;
Best Local Similarity 91.8%; Pred. NO. 4.3e-36;
Matches 89; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

[illegible]

RESULT 3
A87226.
PE-family protein [imported] - Mycobacterium leprae
C/Species: Mycobacterium leprae
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: A87226
R/Coile, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A/Title: Massive gene decay in the leprosy bacillus.
A/Reference number: A86909; MUID:21128732; PMID:11234002
A/Accession: A87226
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-98 <STO>
A/Cross-references: GB:AL450380; NID:gl3093794; PIDN:CAC32063.1; GSPDB:GN000147
C/Genetics:
A/Gene: ML2532

Query Match	76.1%	Score 356	DB 2	Length 98
Best Local Similarity	77.9%	Pred. No. 9.2e-28		
Matches 74	Conservative 7	Mismatches 14	Indels 0	Gaps 0
1	MSLLDAH1POLVASQSAFAAKAGLMRHT1CQAEQAMNSAQAFHQGESSAAFAQAHARFVA	60		
1	MSLLDVH1POLVASESAFAAKAALMRSQ1NQAEEAISAQAFHQGESSAAFAQAHARFVT	60		
61	AAAKVNTLLDVAQANLGEAAAGTVVAADAAAATVY	95		
61	AAAKVNTLLDVAQANLGEAAAGTVVAADAAAATVY	95		

RESULT 4
A70659
hypothetical protein RV2541 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70659
R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70659
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-135 <COL>
A:Cross-references: GB:283863; GB:A1123456; NID:G3261685; PIDN:CAB06197.1; PID:e290876;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2541

RESULT 4
A70659
hypothetical protein RV2541 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70659
R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70659
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-135 <COL>
A:Cross-references: GB:283863; GB:A1123456; NID:G3261685; PIDN:CAB06197.1; PID:e290876;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2541

```

Query Match      19.7%; Score 92; DB 2; Length 135;
Best Local Similarity 36.6%; Pred. NO. 0.047;
Matches 30; Conservative 8; Mismatches 28; Indels 16; Gaps 2;
11 LVASQSAPAKAGLMRHITIGQAEQAAAGSAQAQFHQGESSAAFOAAAHAFVAAAANKVNTLLD 70
59 LAGELSAAVAPSG-----AVLSWQA-----NAVAVNAAHARAGAAAAAVSARMR 102
71 VAQANLGEAAAGTVVAADAAAAA 92
103 ATAAALGCAARRVAGODTAAAA 124

```

71 VAQANLGEAAGTYVAADAAAAS 92
103 ATAAALGOAARRRYAGODTAAAA 124

RESULT 5
H70802 hypothetical protein Rv3874 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70802
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
J. Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70802
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-100 <COL>
A:Cross-references: GB:AL022120; GB:AL123456; NID:G3261558; PIDN:CAA17966.1; PID:g296022
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3874

Query Match	18.7%;	Score 87.5;	DB 2;	Length 100;
Best Local Similarity	34.9%;	Pred. No. 0.095;		
Matches 29;	Conservative 8;	Mismatches 43;	Indels 3;	Gaps 1;
18	FAAKAGLMRHTTIGAAEQAAASQAQAFHOGESAAAFQAAAHAFVFAAAAKVNTLLDVAQANLG	77		
18	PERISGLKTIQIDOVETAGSLQGWGGAAGTAAQAQAVVRFQEAANKQKQELDEISTNIR	77		
78	EAAGTYVAAD--AAAASTVTGF	97		
78	OAGVOYSPRADEEOCALSSQWGF	100		

```

78 EAGTYVAAD---AAAASYTGF 97
      : | | | | | | |
78 OAGVOYSRADDECCOALSSOMGE 100

```

RESULT 6
T34852
Probable secreted protein - Streptomyces coelicolor
C.Species: Streptomyces coelicolor
C.Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C.Accession: T34852
R.Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1999
A.Reference number: Z21559
A.Accession: T34852
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule_type: DNA
A.Residues: 1-1156 <OLI>
A.Cross-references: EMBL:AL035478; PIDN:CAB36606.1; GSPDB:GN00070; SCOEDB:SC2G5.19
A.Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC2G5.19

```
.
Query Match      18.6%; Score 87; DB 2; Length 1156;
Best Local Similarity 37.6%; Pred. No. 1.3;
Matches 35; Conservative 12; Mismatches 34; Indels 12; Gaps 4;

QY          11 LVASQSA--FAAKAGL-----NRHTIGQAEGAAAGAQAFHOGESGAAP-QAAHARFVAAA 62
```

Db 395 LAAQTATCAAAAGISAAATARDASAAAQAQAAVAAQAQSGAAQSEAAVAAAAAADAQA 454
QY 63 AK-----VNTLLDVAQANLGEAAGTYVAADAAA 91
Db 455 ARATKAANRAQSLANTAASAAAAAARKAADSAAA 487

RESULT 7
F70820
hypothetical glycine-rich protein Rv0978c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70820
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70820
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-331 <COL>
A: Cross-references: GB:AL021999; GB:AL123456; NID: g3261538; PIDN: CAAL17577.1; PID: ei25390
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv0978c

Query Match 18.4%; Score 86; DB 2; Length 331;
Best Local Similarity 33.0%; Pred. No. 0.45;
Matches 29; Conservative 14; Mismatches 41; Indels 4; Gaps 2;

QY 9 POLVASQAFAAKAGLMRHTTIGQAEQAQMSQAQAFHQESSAAQFAQ---AHAR-FVAAA 64
Db 8 POLVSTAAADAARIGSAINTANTAAATTVLAAAQDEVSTAAALFGSHGHQYQAISAQ 67

QY 65 VNTLLDVAQANLGEAAGTYVAADAAAAS 92
Db 68 VAAYQORFVLALSQAQSTYVAEAASAT 95

RESULT 8
F70571
hypothetical glycine-rich protein Rv2615c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70571
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70571
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-461 <COL>
A: Cross-references: GB:Z95387; GB:AL123456; NID: g3261763; PIDN: CAB08627.1; PID: g2104298
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv2615c

Query Match 18.4%; Score 86; DB 2; Length 461;
Best Local Similarity 33.0%; Pred. No. 0.64;
Matches 29; Conservative 14; Mismatches 41; Indels 4; Gaps 2;

QY 9 POLVASQAFAAKAGLMRHTTIGQAEQAQMSQAQAFHQESSAAQFAQ---AHAR-FVAAA 64
Db 8 POLVSTAAADAARIGSAINTANTAAATTVLAAAQDEVSTAAALFGSHGHQYQAISAQ 67

QY 65 VNTLLDVAQANLGEAAGTYVAADAAAAS 92
Db 68 VAAYQORFVLALSQAQSTYVAEAASAT 95

RESULT 9
T35781
hypothetical protein SC8A6.14c SC8A6.14c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35781
R: Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A: Reference number: Z21570
A: Accession: T35781
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-1147 <SEE>
A: Cross-references: EMBL:AL031013; PIDN: CAAL19786.1; GSPDB: GNO0070; SCOEDB: SC8A6.14c
A: Experimental source: strain A3(2)
C: Genetics:
A: Gene: SCOEDB: SC8A6.14c

Query Match 18.2%; Score 85; DB 2; Length 1147;
Best Local Similarity 35.1%; Pred. No. 2;
Matches 34; Conservative 11; Mismatches 36; Indels 16; Gaps 3;

QY 12 VASQAFAAKAGLMRHTTIGQAEQAQMS-----QAQFHQESSAAQFAAHARFVAAA 64
Db 312 VTAAANAAAAAGAAASRAAYRAAISAKNANMAEAANQAARVMDVVRTVAAKAK 371

QY 65 VNTLLDVAQANLGEAAGTYV-----AADAASAS 92
Db 372 LAALASDAEAAGNAAAGAAVDSAAAAASADAAAAS 408

RESULT 10
H70820
hypothetical glycine-rich protein Rv0980c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70820
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: H70820
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-457 <COL>
A: Cross-references: GB:AL021999; GB:AL123456; NID: g3261538; PIDN: CAAL17579.1; PID: g2916941
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv0980c

Query Match 17.9%; Score 84; DB 2; Length 457;
Best Local Similarity 31.8%; Pred. No. 0.99;
Matches 28; Conservative 16; Mismatches 40; Indels 4; Gaps 2;

QY 9 POLVASQAFAAKAGLMRHTTIGQAEQAQMSQAQAFHQESSAAQFAQ---AHAR-FVAAA 64
Db 8 POLVSTAAADAARIGSAINTANTAAATTVLAAAQDEVSTAAALFGSHGHQYQAISAQ 67

QY 65 VNTLLDVAQANLGEAAGTYVAADAAAAS 92
Db 68 VAAYQORFVLALSQAQSTYVAEAASAT 95

RESULT 11

```
G70682
hypothetical glycine-rich protein RV2396 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: G70682
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406933; PMID:10952301
A:Accession: F82181
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-405 <HEI>
A:Cross-references: GB:AE004236; GB:AE003852; NID:99656095; PIDN:AAFP94749.1; GSPDB:GNC011
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: RV2396
Query Match 17.2%; Score 80.5; DB 2; Length 361;
Best Local Similarity 35.1%; Pred. No. 1.7; Mismatches 37; Indels 13; Gaps 3;
Matches 33; Conservative 11;
QY 11 LVASQSAFAAKAGLMRHRTIGQAEQAAMSAQAQFHQGE-----SSAFAQAAHAR-F 58
Db 4 LIASPEALANTATYLT-CIGSAISANAVAAAPTTEILLAGTDEVSTAISALFCAHAQAY 62
QY 59 VAAAKVNTLLDVAQNLGEAGTIVAAADAAAS 92
Db 63 QALSAAHVAAFHQDFVHTLTAGAGSYMAAEAAAS 96
RESULT 12
G70881
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70881
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70881
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <COL>
A:Cross-references: GB:AL008967; GB:AL123456; NID:93261491; PIDN:CAA15564.1; PID:el17389
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE
Query Match 17.2%; Score 80.5; DB 2; Length 394;
Best Local Similarity 36.7%; Pred. No. 1.9; Mismatches 37; Indels 7; Gaps 5;
Matches 33; Conservative 13;
QY 9 POLVASQSAFAAKAGLMRHRTIGQAEQAAMSAQAQFHQGE---SSAFAQAAHARFVAAAKV 65
Db 107 PRAEVANRALLA-ALVATNVLCQNTPAIMATEA-HYGENWAGDALMYGYAASGAAGRL 164
QY 66 NTLILDVAQ-ANLGEAGTIVA-ADAAAST 93
Db 165 NPLITPSQTANVAGLAGQAQAAVSHAAAST 194
RESULT 13
F82181
galactokinase VC1595 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70504
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82181
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406933; PMID:10952301
A:Accession: F82181
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-405 <HEI>
A:Cross-references: GB:AE004236; GB:AE003852; NID:99656095; PIDN:AAFP94749.1; GSPDB:GNC011
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1595
A:Map position: 1
A:Superfamily: galactokinase
Query Match 17.1%; Score 80; DB 2; Length 405;
Best Local Similarity 27.3%; Pred. No. 2.1; Mismatches 39; Indels 8; Gaps 3;
Matches 24; Conservative 17;
QY 2 SLDDAHIPOLVASQSAF-AAKAGLMRHRTIGQAEQAAMSAQAQFHQGES---SAAFAQAAHAR 57
Db 261 ALRDVTYLAQLTAKQAEIDPVVAKBPAHVITENRTLHAAQALREGMPLGELMAASHAS 320
QY 58 ----FVAAAANKVNTLLDVAQNLGEAG 81
Db 321 MRDDFEITVKEIDTLVEIVQSVIGDQGG 348
RESULT 14
G70929
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:93250699; PIDN:CAA17711.1; PID:el25460
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE
Query Match 16.8%; Score 77.5; DB 2; Length 393;
Best Local Similarity 31.4%; Pred. No. 3.6; Mismatches 41; Indels 19; Gaps 4;
Matches 32; Conservative 10;
QY 9 POLVASQSAFAAKAGLMRHRTIGQAEQAAMSAQAQFHQGESAAFAQAAHARFV---AAAAKV 65
Db 60 PASAANAEAVAPVAVNWSAAQAQAEQAATCARA----AAAAFEAAFAATVPPLIAANR 114
QY 66 NTLILDVAQNL-GEAGTIVAA-----DAAASTYTG 96
Db 115 ASLQMLISTNVFGNTSAIAAAEAQYGENWAGDAAAMAYAG 156
RESULT 15
A70504
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70504
```


Search completed: June 23, 2004, 17:13:38
Job time : 27 secs

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 17 Seconds
(without alignments)
297.106 Million cell updates/sec

Title: US-09-886-349A-24

Perfect score: 468

Sequence: 1 MSLLDAHLPQLVASQSAFAA.....EAGGTVAADAAAATYGTG 97

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	87.5	18.7	99	1 ESXB_MYCTU	O69739 mycobacteri
2	80	17.1	386	1 GALL_VTBCH	O9kxp1 vibrio chol
3	77.5	16.6	1238	1 SBCC_RHOCA	O68032 rhodobacter
4	75.5	16.1	428	1 FXB2_MOUSE	O64733 mus musculu
5	75.5	16.1	592	1 CEA_CITPR	P04480 citrobacter
6	75	16.0	339	1 PF5B_LOLPR	Q40237 lolium pete
7	74.5	15.9	222	1 DEOC_CORGL	O8ntc4 corynebacte
8	74.5	15.9	436	1 Y868_CHLMU	O9p1g1 chlamydia m
9	74.5	15.9	498	1 Y118_MYCTU	O50615 mycobacteri
10	74	15.8	97	1 ANP_LIMEP	P09031 limanda fer
11	74	15.8	774	1 STPL_LAMED	P03764 bacterioph
12	74	15.8	880	1 BRC4_DROME	Q24206 drosophila
13	73.5	15.7	558	1 YJ83_MYCTU	Q10873 mycobacteri
14	73	15.6	357	1 RF1_MYCTU	Q10605 mycobacteri
15	72.5	15.5	668	1 PAU_DROME	O9vgx3 drosophila
16	72.5	15.5	801	1 Y747_MYCTU	O53810 mycobacteri
17	72.5	15.5	957	1 Y278_MYCTU	P56877 mycobacteri
18	72.5	15.5	1023	1 RT12_ACTPL	P55129 actinobacil
19	72	15.4	905	1 CTFP_MYCTU	Q10860 mycobacteri
20	71.5	15.3	263	1 PSP1_ARATH	Q42029 arabidopsis
21	71	15.2	324	1 CSA_DROSI	Q9n6k2 drosophila
22	70.5	15.1	163	1 U338_LOOMI	P04375 locusta mig
23	69.5	14.9	274	1 PYRF_MYCTU	P77838 mycobacteri
24	69	14.7	276	1 PYRF_MYCTU	O08323 mycobacteri
25	69	14.7	395	1 TBLU_AGRU	P34913 agrobacteri
26	69	14.7	408	1 YS92_MYCTU	Q10813 mycobacteri
27	69	14.7	474	1 CYAE_BORPE	P11032 bordetella
28	68.5	14.6	102	1 YD86_MYCTU	P71656 mycobacteri
29	68.5	14.6	275	1 HUPH_RHOCA	Q03007 rhodobacter
30	68.5	14.6	2541	1 TLN1_HUMAN	O9y490 homo sapien
31	68.5	14.6	2541	1 TLN1_MOUSE	P26039 mus musculu
32	68	14.5	866	1 AREA_ASPOR	O33415 aspergillus
33	67.5	14.4	678	1 YF48_MYCTU	Q10778 mycobacteri

34	67.5	14.4	1023	1 RT11_ACTPL	P55128 actinobacil
35	67	14.3	276	1 SX21_HUMAN	O9y451 homo sapien
36	67	14.3	2716	1 OSA_DROME	Q8in94 drosophila
37	66.5	14.2	355	1 YH11_ECOLI	P37626 escherichia
38	66.5	14.2	360	1 ALF_DROME	P07764 drosophila
39	66.5	14.2	415	1 FOX1_CAEEL	Q10572 caenorhabdi
40	66.5	14.2	550	1 CCF_DROME	P41046 drosophila
41	66.5	14.2	697	1 SIM_DROME	P05709 drosophila
42	66.5	14.2	1120	1 STFR_ECOLI	P76072 escherichia
43	66	14.1	176	1 VP26_EBV	PL4348 Epstein-Bar
44	66	14.1	441	1 ZRAR_ECO57	Q8x613 escherichia
45	66	14.1	511	1 GPM1_STRCO	Q91214 streptomyc

ALIGNMENTS

RESULT 1

ESXB_MYCTU

ID ESXB_MYCTU STANDARD; PRT; 99 AA.

AC O69739;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ESAT-6 like protein esxB (10 kDa culture filtrate antigen cfp10)
DE (Secreted antigenic protein MTS-10).
GN ESXB OR CFP10 OR LHP OR RV3874 OR MT3988 OR MTV027.09 OR
GN MB3904.

OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=99061212; PubMed=9846755;
RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
RA Gicquel B.;

RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
RT low-molecular-mass culture filtrate protein (CFP-10).";
RL Microbiology 144:3195-3203 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Flietschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;

RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis;
RX Singh B., Siddiqui Z., Singh S., Sharma P.;

[illegible]

```

RESULT 3
ID          SBCC_RHOCA
AAC         SBCC_RHOCA
AAC         STANDARD;          PRT; 1238 AA.
ID          C68032;
AAC         28-FEB-2003 (Rel. 41, Created)
AAC         28-FEB-2003 (Rel. 41, Last sequence update)
AAC         28-FEB-2003 (Rel. 41, Last annotation update)
AAC         Nuclease sbccd subunit C.
AAC         SBCC.
AAC         Rhodobacter capsulatus (Rhodopseudomonas capsulata).
AAC         Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
AAC         Rhodobacteraceae; Rhodobacter.
AAC         NCBI_TaxID=1061;
AAC         [1]
AAC         SEQUENCE FROM N.A.
AAC         STRAIN=SB1003 / St Louis;
AAC         MEDLINE=97404404; PubMed=9256491;
AAC         Vleck C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
AAC         "Sequence of a 189-kb segment of the chromosome of Rhodobacter
AAC         capsulatus SB1003.";
AAC         Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
AAC         -!- FUNCTION: SBCC cleaves DNA hairpin structures. These structures
AAC         can inhibit DNA replication.

```



```
QY 63 AKYNTLLDVAQANLGEAGTYVAADAAAAS 93
Db 300 ATATATPAAAYATATPAAATATATPAAATAT 330

RESULT 7
DEOC_CORGL STANDARD; PRT; 222 AA.
AC QNTG4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2001 (Rel. 40, Last annotation update)
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)
DE (deoxyriboaldolase) (DERA).
GN DEOC OR GLO383.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-
CC glyceraldehyde 3-phosphate + acetaldehyde.
CC -I- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: Belongs to the deoc/fbaB aldolase family. Deoc
CC subfamily 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP005275; BAB97776.1; -.
CC HAMAP; MF_00114; -.
CC InterPro; IPR002915; Deoc.
CC InterPro; IPR003009; FMN_enzyme.
CC Pfam; PF01791; Deoc; 1.
CC Lysase; Schiff base; Complete proteome.
FT BINDING 156 156 SCHIFF-BASE WITH ACETALDEHYDE (BY
FT SIMILARITY).
FT ACT_SITE 186 186 BY SIMILARITY.
SQ SEQUENCE 222 AA; 22520 MW; 563C3D6B6805DD08 CRC64;

Query Match 15.9%; Score 74.5; DB 1; Length 222;
Best Local Similarity 21.7%; Pred. No. 1.9;
Matches 28; Conservative 18; Mismatches 34; Indels 49; Gaps 3;

QY 8 IPQLVASQSFAKAGLMRHTIGAEQAAMSAQAFHGV--AAAKV 65
Db 43 VPNSMWNLTAKAQAGIRVATVA-----GFPHGKTPALVKAAEARLAVQSGASEV 92

QY 66 NTLLDVAQANLGE-----AACTYVAADA 88
Db 93 DVVLIDIAVVEGDNRLLQEIIVAIREAVPSVULKFILETAVVSDIAIVTAVNALIAGA 152

QY 89 AAASVTYGF 97
Db 153 DFKATSTGF 161

RESULT 8
Y868_CHLMU
ID Y868_CHLMU STANDARD; PRT; 436 AA.
AC Q9FJG1;
DT 16-OCT-2001 (Rel. 40, Created)

QY 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TC0868.
GN TC0868.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F., Bass S.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Neilson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.W.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -I- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0808/CT579/TC0868
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002353; AAF39664.1; -.
CC PIR; G81655; G81655.
CC TIGR; TC0868; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 436 AA; 43891 MW; 43E21C9924CD7A48 CRC64;

Query Match 15.9%; Score 74.5; DB 1; Length 436;
Best Local Similarity 28.2%; Pred. No. 3.6;
Matches 33; Conservative 10; Mismatches 39; Indels 35; Gaps 3;

QY 10 QLVASQSFAKAGLMRHTIGAEQAAMSAQAFHQ----- 44
Db 163 QAIQCVMAPEIGNAIRTOHQHQAATELQAKQSLISGINTIVGFVAVSGGILSASKS 222

QY 45 --GESSAAF-----QAAHAFVAAAKVNTLLDVAQANLGEAGTYVAADAAAAS 93
Db 223 LGGLKSAFTNETATATASSATSAAKTAVNALDDV--ANVAATAGTKAASGAASAAS 277

RESULT 9
YI18_MYCTU
ID YI18_MYCTU STANDARD; PRT; 498 AA.
AC Q50615;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PE-PERS family protein RV1818C/MT1866.
GN RV1818C OR MT1866 OR MTCY1A11.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellton S., Squares S., Squares R.,
```

```

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Ginn M., Haft D., Kickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J.A., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGKS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z8020; CAB01461.1; -.
DR EMBL; AE007045; AAK46139.1; ALT_INIT.
DR PIR; C70720; C70720.
DR TIGR; MT1866; -.
DR Tuberculist; Rv1818c; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
DR Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 403 423 POTENTIAL.
FT TRANSMEM 414 414 MISSING (IN REF. 2).
FT CONFLICT 183 212 G -> GAGG (IN REF. 2).
FT CONFLICT 414 414
SQ SEQUENCE 498 AA; 40755 MW; 4F6F78F2482586BA CRC64;

Query Match 15.9%; Score 74.5; DB 1; Length 498;
Best Local Similarity 25.9%; Pred. No. 4.1;
Matches 29; Conservative 17; Mismatches 40; Indels 11; Gaps 3;

QY 8 IPVLVQSQAFAAKAGLMRTHTIGQAEQAA-----MSAQAFHQGSSAAFOAHAR-FV 59
DB 7 IPEALAAVATDLAGIG---STIGTANAAAAPVPTTVLAAAADBSVSAAMALFSGHAQAYQ 63
QY 60 AAAAKVNTLIDVAQANLGEAGTYVAAADAAAASYTG 96
DB 64 ALSAQALFHEQFVRLTAGAGSYAAAEASAAPLEG 100

RESULT 10
ANP LIMFE STANDARD; PRT; 97 AA.
AC P05031;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)

STP_LAMBD STP_LAMBD STANDARD; PRT; 774 AA.
ID STP_LAMBD STANDARD; PRT; 774 AA.
AC P03764; P03745;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Side tail fiber protein.
GN SIF.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=10710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA.";
RL J. Mol. Biol. 162:729-773(1982).
RN [2]
RP IDENTIFICATION AS STF.
RX MEDLINE=92165720; PubMed=1531648;

```


RA Haggaard-Ljungquist E., Halling C., Calendar R.;
 RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence
 RT for horizontal transfer of tail fiber genes among unrelated
 RT bacteriophages.";
 RL J. Bacteriol. 174:1462-1477(1992).
 RN [3]
 RP RECONSTRUCTION.
 RA MEDLINE=93068310; PubMed=1439823;
 RX Hendrix R.W., Duda R.L.;
 RL "Bacteriophage lambda PaPa: not the mother of all lambda phages.";
 RL Science 258:1145-1148(1992).
 CC -1- MISCELLANEOUS: The common laboratory strain of bacteriophage
 CC lambda; the original isolate. The Ur-lambda virions have thin,
 CC lambda; the original isolate. The Ur-lambda virions have thin,
 CC joined tail fibers (side tail fibers) that are absent from lambda
 CC wild type. Relative to lambda PaPa, Ur-lambda has expanded
 CC receptor specificity and adsorbs to E.coli cells more rapidly.
 CC -1- SIMILARITY: Belongs to the tail fiber family.
 CC -1- CAUTION: This is a conceptual translation; a frameshift was
 CC corrected in position to 396 to recreate the original stf protein.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement.(See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J02459; AAA96555.1; ALT FRAME.
 DR EMBL; J02459; AAA96557.1; ALT FRAME.
 DR PIR; C43009; QXBP2L.
 DR InterPro; IPR008969; Carboxypeptidase.
 DR InterPro; IPR005003; Phage_fiber.
 DR InterPro; IPR005068; Phage_fiber_2.
 DR Pfam; PF03335; Phage_fiber; 6.
 DR Pfam; PF03406; Phage_fiber_2; 1.
 KW Fiber protein.
 SQ SEQUENCE 774 AA; 77527 MW; CDD1DF85E919123B CRC64;
 Query Match 15.8%; Score 74; DB 1; Length 774;
 Best Local Similarity 32.2%; Pred. No. 6.8;
 Matches 29; Conservative 11; Mismatches 40; Indels 10; Gaps 2;
 QY 6 AHIPQLV--ASQSAFAAKAGLMRHHTIGQAQAAMSAQAQFHGESSAAFOAAHARFVAAAA 63
 Db 139 AQVALVTDTDSARAAT-----SAGQAASSQAQSSGAEASAKATEAKSAAAA 190
 QY 64 KVTLLDVAQNLGEAAGTVVADAAAST 93
 Db 191 ESSKNAATSAGAAKTSETNAAASQQAAT 220
 RESULT 12
 BRCA_DROME STANDARD; PRT; 880 AA.
 AC Q24206; Q46064; Q9W575;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Broad-complex core-protein isoform 6.
 GN BR OR BR-C OR EG:17A9.1 OR EG:25D2.1 OR EG:123F11.1 OR
 GN CG11491/CG11514.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND CHARACTERIZATION OF
 RP ISOFORMS.
 RC TISSUE=Imaginal disks, and Larva;
 RX MEDLINE=96299417; PubMed=8660872;

RA Bayer C.A., Holley B., Fristrom J.W.;
 RT "A switch in broad-complex zinc-finger isoform expression is regulated
 RT posttranscriptionally during the metamorphosis of Drosophila imaginal
 RT discs.";
 RL Dev. Biol. 177:1-14(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D., Cadieu E.,
 RA Bartell B.G., Ferraz C., Vidal S., Brun C., Denaillies J., Borkova D.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo S.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourikiofi F.,
 RA Beiernt N., Dowe G., Schaefer U., Jaesckle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamitso A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of D.
 RT melanogaster.";
 RL Science 287:2220-2222(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceinkner S.B., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin L.,
 RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler K.J., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gloeck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalw C.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP CHARACTERIZATION OF ISOFORMS, AND MUTATIONAL ANALYSIS.
 RX MEDLINE=97384928; PubMed=9242423;
 RA Bayer C.A., von Kalm L., Fristrom J.W.;
 RT "Relationships between protein isoforms and genetic functions
 RT demonstrate functional redundancy at the Broad-Complex during
 RT Drosophila metamorphosis.";
 RL Dev. Biol. 187:267-282(1997).
 CC -1- FUNCTION: BROAD-COMPLEX PROTEINS ARE REQUIRED FOR PUFFING AND

KW	Hypothetical protein; Transmembrane; Complete proteome.	
FT	DOMAIN 27 53 ALA-RICH.	
FT	TRANSMEM 332 352 POTENTIAL.	
FT	TRANSMEM 360 380 POTENTIAL.	
SEQ	SEQUENCE 558 AA; 53738 MW; 17ECBE43778E021E CRC64;	
Query Match	15.7%; Score 73.5; DB 1; Length 558;	
Best Local Similarity	30.0%; Pred. No. 5.6; Indels 5; Gaps 2;	
Matches	27; Conservative 12; Mismatches 46; Indels 5; Gaps 2;	
Qy	8 IPOLVASQFAAKAGLMRHTTICQAEQAMSA-QAFHOGCESSAAFAQAAHARF-----VAAA 62	
Db	7 VPEFLTSAADVENIGSTLRAANAASASTTALAAGADEVSAAVAALFAFFGQYQAVS 66	
Qy	63 AKNTLLDVAQANLGEAGTYVAADAAAAS 92	
Db	67 AQASAFHQPFVQTLNSAGSYAAAEATIAS 96	
RESULT 14		
RF1_MYCTU	STANDARD; PRT; 357 AA.	
ID	RF1_MYCTU	
AC	Q10605;	
DT	01-OCT-1996 (Rel. 34, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Peptide chain release factor 1 (RP-1).	
GN	PRFA OR RV1299 OR MT1338 OR MTCY373.19 OR MB1311.	
OS	Mycobacterium tuberculosis, and	
OS	Mycobacterium bovis.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773, 1765;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	SPECIES=M.tuberculosis; STRAIN=H37RV;	
RX	MEDLINE=98295987; PubMed=9634230;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.B. III, Tekaita F.,	
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,	
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,	
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;	
RT	"Deciphering the biology of Mycobacterium tuberculosis from the	
RT	complete genome sequence.";	
RL	Nature 393:537-544 (1998).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;	
RX	MEDLINE=22206494; PubMed=12218036;	
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,	
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,	
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,	
RA	Deicher A., Usterback T., Weidman J., Khouri H., Gill J., Mikula A.,	
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;	
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and	
RT	laboratory strains.";	
RL	J. Bacteriol. 184:5479-5490 (2002).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	SPECIES=M.bovis; STRAIN=AF2122/97;	
RX	MEDLINE=22709107; PubMed=12788972;	
RA	Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,	
RA	Pryor M., Duthoy S., Grondin S., Lacroix C., Monsepe C., Simon S.,	
RA	Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,	
RA	Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;	
RT	"The complete genome sequence of Mycobacterium bovis.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).	
CC	-!- FUNCTION: Peptide chain release factor 1 directs the termination	
CC	of translation in response to the peptide chain termination codons	
CC	UAG and UAA (By similarity).	

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 39 Seconds
(without alignments)
784.750 Million cell updates/sec

Title: US-09-886-349A-24

Perfect score: 468
Sequence: 1 MSLLDAHIPQLVASQSAFAA.....EAAGTYVAADAAASTYTGTF 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPREMBL 25:**

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	100.0	97	16	O53692 mycobacteri
2	468	100.0	97	16	Q7U2C8 mycobacteri
3	442	94.4	97	16	O53267 mycobacteri
4	442	94.4	102	16	Q7TXF7 mycobacteri
5	442	94.4	102	16	Q8VJ87 mycobacteri
6	356	76.1	98	16	Q9CD32 mycobacteri
7	92	19.7	135	16	P95012 mycobacteri
8	92	19.7	135	16	Q7TYC5 mycobacteri
9	92	19.7	273	16	Q8VJG4 mycobacteri
10	89	19.0	457	16	Q7U0X5 mycobacteri
11	87	18.6	1156	16	Q9Z5A4 streptomyc
12	86	18.4	331	16	O53891 mycobacteri
13	86	18.4	335	16	Q8VKA6 mycobacteri
14	86	18.4	476	16	O06199 mycobacteri
15	85	18.2	1147	16	O87848 streptomyc
16	84	17.9	335	16	Q7U0X7 mycobacteri

17	84	17.9	457	16	O53693
18	84	17.9	464	16	Q7TY83 mycobacteri
19	84	17.9	476	16	Q8VKA5 mycobacteri
20	80.5	17.2	382	16	P71748 mycobacteri
21	80.5	17.2	394	16	Q7TXK5 mycobacteri
22	80.5	17.2	462	16	O33310 mycobacteri
23	79.5	17.0	92	13	Q9DF23 myoxocephal
24	79.5	17.0	3703	5	Q8IMW2 drosophila
25	79.5	17.0	3912	5	Q9VC56 drosophila
26	79	16.9	113	16	Q8PMS9 xanthomonas
27	79	16.9	113	16	Q8PB30 xanthomonas
28	77.5	16.6	393	16	Q7TJ23 mycobacteri
29	77.5	16.6	394	16	O33205 mycobacteri
30	77.5	16.6	394	16	Q7TTR6 mycobacteri
31	77.5	16.6	411	16	O53939 mycobacteri
32	77.5	16.6	956	16	Q8PGH7 xanthomonas
33	77	16.5	459	16	Q88819 pseudomonas
34	77	16.5	753	2	Q845W1 burkholderi
35	77	16.5	951	5	Q9VEC6 drosophila
36	77	16.5	984	5	Q8IN89 drosophila
37	77	16.5	1366	16	Q8Y120 raistonia s
38	76.5	16.3	107	10	Q9FEY2 heterocapsa
39	76.5	16.3	172	16	Q8VKF1 mycobacteri
40	76.5	16.3	172	16	Q7UID8 mycobacteri
41	76.5	16.3	175	16	O53805 mycobacteri
42	76.5	16.3	201	10	Q9FEY3 heterocapsa
43	76.5	16.3	354	2	Q9AEZ5 frankia sp.
44	76.5	16.3	790	16	Q8ZQF5 salmonella
45	76	16.2	376	16	Q8Z8C1 salmonella

ALIGNMENTS

RESULT 1

O53692 PRELIMINARY; PRT; 97 AA.
ID US3692
AC O53692;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DE Hypothetical protein (PE family protein).
GN RV0287 OR MTW035.15 OR MT0300.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RI complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

```

DR  EMBL; AL021930; CAA17362.1; -.
DR  EMBL; AE006937; AAK44524.1; -.
DR  PIR; E70836; E70836.
DR  TIGR; MT0300; -.
DR  DR  Hypothetical protein; Complete proteome.
KW  SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;

  Query Match      100.0%; Score 468; DB 16; Length 97;
  Best Local Similarity 100.0%; Pred. No. 3.3e-37;
  Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60
Db  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60

Qy  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97
Db  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97

RESULT 2
Q7U2C8      PRELIMINARY;      PRT;      97 AA.
ID  Q7U2C8
AC  Q7U2C8; PubMed=12788972;
DT  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB  Conserved hypothetical protein TB9.8.
GN  ESXG OR MB0255
OS  Mycobacterium bovis.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX  NCBI_TaxID=1765;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AF2122/97;
RX  MEDLINE=22709107; PubMed=12788972;
RA  Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA  Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA  Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA  Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA  "The complete genome sequence of Mycobacterium bovis.";
RT  Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL  EMBL; BX248335; CAD93159.1; -.
DR  EMBL; BX248335; CAD93159.1; -.
KW  Complete proteome.
SQ  SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;

  Query Match      100.0%; Score 468; DB 16; Length 97;
  Best Local Similarity 100.0%; Pred. No. 3.3e-37;
  Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60
Db  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60

Qy  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97
Db  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97

RESULT 3
O53267      PRELIMINARY;      PRT;      97 AA.
ID  O53267
AC  O53267; PubMed=12788972;
DT  01-JUN-1998 (TrEMBLrel. 06, Created)
DT  01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DB  PE-family protein.
GN  RV3020C OR MTV012.34C.
OS  Mycobacterium tuberculosis.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Mycobacteriaceae; Mycobacterium.

  Query Match      100.0%; Score 468; DB 16; Length 97;
  Best Local Similarity 100.0%; Pred. No. 3.3e-37;
  Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60
Db  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60

Qy  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97
Db  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97

RESULT 4
O7TXF7      PRELIMINARY;      PRT;      97 AA.
ID  O7TXF7
AC  O7TXF7; PubMed=12788972;
DT  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB  PE family protein.
GN  ESXS OR MB3046C.
OS  Mycobacterium bovis.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX  NCBI_TaxID=1765;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AF2122/97;
RX  MEDLINE=22709107; PubMed=12788972;
RA  Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA  Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA  Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA  Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA  "The complete genome sequence of Mycobacterium bovis.";
RT  Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL  EMBL; BX248344; CAD96733.1; -.
DR  EMBL; BX248344; CAD96733.1; -.
KW  Complete proteome.
SQ  SEQUENCE 97 AA; 9842 MW; BA9BCB3180EC17F2 CRC64;

  Query Match      94.4%; Score 442; DB 16; Length 97;
  Best Local Similarity 91.8%; Pred. No. 9.8e-35;
  Matches 89; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60
Db  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60

Qy  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97
Db  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97

  Query Match      94.4%; Score 442; DB 16; Length 97;
  Best Local Similarity 91.8%; Pred. No. 9.8e-35;
  Matches 89; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60
Db  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAQFHQGSAAFAQAAHARFVA 60

Qy  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97
Db  61 AAANKVNTLLDVAQNGLGEAGTYVAADAAAASYTGF 97

```

```
RESULT 5
Q8VJ87
ID Q8VJ87 PRELIMINARY; PRT; 102 AA.
AC Q8VJ87;
DT 01-NAR-2002 (TReMBLrel. 20, Created)
DT 01-NAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PE family protein.
GN MT3105.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007129; AAK47434.1; -.
DR TIGR; MT3105; -.
SQ SEQUENCE 102 AA; 10372 MW; 03CB984D6633D2E4 CRC64;

Query Match 94.4%; Score 442; DB 16; Length 102;
Best Local Similarity 91.8%; Pred. No. 1e-34; 3; Indels 0; Gaps 0;
Matches 89; Conservative

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFAQAAHARFVA 60
Db 6 MSLLDAHIPOLVASHAFPAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFAQAAHARFVA 65

QY 61 AAKVNTLLDVAQANLGEAAGTYVAADAAASTYTG 97
Db 66 AAKVNTLLDVAQANLGEAAGTYVAADAAASTYTG 102

RESULT 6
Q9CD32
ID Q9CD32 PRELIMINARY; PRT; 98 AA.
AC Q9CD32;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PE-family protein.
GN M2532.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1769;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=IN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeter K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Halroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RL "Massive gene decay in the leprosy bacillus.";
RT Nature 409:1007-1011 (2001).
DR EMBL; AL583926; CAC32063.1; -.
DR FIR; A87226; A87226.

Query Match 19.7%; Score 92; DB 16; Length 135;
Best Local Similarity 36.6%; Pred. No. 0.29;
Matches 30; Conservative 8; Mismatches 28; Indels 16; Gaps 2;

QY 11 LVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFAQAAHARFVAANKVNTLLD 70
Db 59 LAGLSAAVAPSG-----AVLSWQA-----NAVAVNAAHARAGAAAAVSAEMR 102

QY 71 VAQANLGEAAGTYVAADAAAS 92
Db 103 ATAAALGOARRYAGQDTAAAA 124

RESULT 8
Q7TYC5
ID Q7TYC5 PRELIMINARY; PRT; 135 AA.
AC Q7TYC5;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical alanine rich protein.
GN MB2370.
```

	Query Match	19.7%	Score 92;	DB 16;	Length 373;
	Best Local Similarity	36.6%;	Pred.No.	0.93;	
	Matches 30; Conservative	8;	Mismatches	28;	Indels 16; Gaps 2;
QY	11 LVASQSAPAKAGLMRHITIGAECAWMSAQAFHQGESAAFOAHARTVAAAAXKNTLLD	70	:	:	:
Dd	59 LAGELSAARVPSG-----AVLSMQA-----NAVAVNNAHARAAGAAAAVA\$MR	102	:	:	:
QY	71 VAQNGLGEAGTYTVAADAAAAS	92	:	:	:
Dd	103 ATAAALGOARRRYAGODTAATAA	124	:	:	:

RESULT	11
ID	Q9ZSA4
AC	PRELIMINARY;
DC	Q9ZSA4
CC	Q9ZSA4;
DT	01-MAY-1999 (TrEMBLrel. 10, Created)
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Putative secreted protein.
GN	SCO6198 OR SC3G5.19
GE	Streptomyces coelicolor.
OS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC	Streptomycineae; Streptomycetaceae; Streptomyces.
OX	NCBI_TaxID=1902;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=A3(2);
RA	Oliver K.; Harris D.;
RL	Submitted (FEB-1999) to the EMBL/GenBank/DDBB databases.

```

RESULT 11
Q9ZSA4 PRELIMINARY; PRT; 1156 AA.
ID Q9ZSA4;
AC Q9ZSA4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative secreted protein.
OS SCO6198 OR SC235.19.
OC Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteriales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K.; Harris D.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D.; Parkhill J.; Barrell B.G.; Rajandream M.A.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
MEDLINE=970000351; PubMed=8843436;
RA Redenbach M.; Kisser H.M.; Denapite D.; Eichner A.; Cullum J.;
RA Kinashi H.; Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RN [4]

```



```
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzarrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- COFACTOR: ZINC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL; AL939126; CAB36606.1; -.
DR PIR; T34852; T34852.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR002328; ADH zinc.
DR InterPro; IPR005506; DUF312.
DR Pfam; PF03752; ALF; 8.
DR PROSITE; PS0059; ADH_ZINC; 1.
DR Oxidoreductase; Zinc; Complete proteome.
KW SEQUENCE 1156 AA; 116717 MW; 7046B814FB9517F8 CRC64;

Query Match 18.6%; Score 87; DB 16; Length 1156;
Best Local Similarity 37.6%; Pred. No. 10;
Matches 35; Conservative 12; Mismatches 34; Indels 12; Gaps 4;

QY 11 LVASQSA--FAAKAGL-----NRHTIGQAEQAAMSAQAAPHQGESSAAAF-QAAHARFVAAA 62
DB 395 LAAAQATQCAAAAGISAAATARDSSAAQAQAAVAQAASGAQAQSAVAARAAABADAQA 454

QY 63 AK-----VNTLLDVAQANLGEAAGTYVAADAAA 91
DB 455 ARATKAANRAQSLANTAAASAAAARKAADSA 487

RESULT 12
Q53891 PRELIMINARY; PRT; 331 AA.
AC Q53891
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DE PGSR-family protein.
GN RV0978C OR MTW044.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; AL021999; CAA17577.1; -.
DR PIR; F70820; F70820.
```

```
DR TubercuList; RV0978C; -.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF01436; NHL; 3.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Complete proteome.
SQ SEQUENCE 331 AA; 31019 MW; A093EC12F1691CA5 CRC64;

Query Match 18.4%; Score 86; DB 16; Length 331;
Best Local Similarity 33.0%; Pred. No. 3;
Matches 29; Conservative 14; Mismatches 41; Indels 4; Gaps 2;

QY 9 POLVASQSAFAAKAGLNRHTIGQAEQAAMSAQAAPHQGESSAAFOA---AHAR-FVAAA 64
DB 8 POLVSTAAADAARIGSAINTANTAAATTTQVLAQAQDEVSTAALFGSHGQHYQAISAQ 67

QY 65 VNTLLDVAQANLGEAAGTYVAADAAAAS 92
DB 68 VAAVQQRFLVLSQAGSTYAVAEASAT 95

RESULT 13
Q8VKAG PRELIMINARY; PRT; 335 AA.
AC Q8VKAG
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE PE_PGSR family protein.
GN MT1006.1.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Deicher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006985; AAK45254.1; -.
DR TIGR; MT1006; -.
DR InterPro; IPR001258; NHL.
DR Pfam; PF01436; NHL; 3.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
SQ SEQUENCE 335 AA; 31351 MW; C2C742369C499244 CRC64;

Query Match 18.4%; Score 86; DB 16; Length 335;
Best Local Similarity 33.0%; Pred. No. 3.1;
Matches 29; Conservative 14; Mismatches 41; Indels 4; Gaps 2;

QY 9 POLVASQSAFAAKAGLNRHTIGQAEQAAMSAQAAPHQGESSAAFOA---AHAR-FVAAA 64
DB 8 POLVSTAAADAARIGSAINTANTAAATTTQVLAQAQDEVSTAALFGSHGQHYQAISAQ 67

QY 65 VNTLLDVAQANLGEAAGTYVAADAAAAS 92
DB 68 VAAVQQRFLVLSQAGSTYAVAEASAT 95

RESULT 14
Q06199 PRELIMINARY; PRT; 476 AA.
AC Q06199
DT 01-JUL-1997 (TREMBlrel. 04, Created)
```

